TỦ SÁCH SPUTNIK

Sách điện tử SE001

GS. Nguyễn Tiến Dũng và GS. Đỗ Đức Thái



NHẬP MÔN HIỆN ĐẠI
XÁC SUẤT & THỐNG KÊ



Đây là phiên bản điện tử miễn phí
dành cho các bạn đọc của
Sputnik Education

Phiên bản này: Ngày 3 tháng 5 năm 2015

Mục lục

Lời tựa cho bản e-book

Cuốn sách này được in ra lần đầu vào năm 2010. Các tác giả đã bỏ rất nhiều tâm trí và sức lực để viết nó, nhằm đạt chất lượng tốt nhất có thể. Mong muốn của các tác giả là làm sao cuốn sách được phổ biến thật rộng rãi ở Việt Nam, đặc biệt là ở các trường đại học, để giúp các bạn sinh viên tiếp cận được với xác suất thống kê một cách dễ hiểu hơn, đúng bản chất hơn, dễ ứng dụng hơn.

Từ lúc in ra năm 2010, cuốn sách đã nhận được rất nhiều phản hồi tích cực từ phía bạn đọc về mặt nội dung. Về mặt chất lượng in ấn và phát hành thì không được tốt bằng, và rất tiếc những khâu đó nằm ngoài khả năng kiểm soát của các tác giả. Hiện tại bản in năm 2010 không còn trên thị trường, và các tác giả nhận được thư của hàng trăm người nói rằng muốn sách tái bản.

Để có thể phục vụ tốt hơn các bạn đọc, đặc biệt là các bạn sinh viên, các tác giả đã kết hợp với **Tủ Sách Sputnik** công bố miễn phí bản điện tử của cuốn sách này. Một số lỗi trong bản in năm 2010 đã được sửa trong bản điện tử này.

Tủ Sách Sputnik của Sputnik Education, mà các tác giả tham gia

làm cộng tác viên, là một dự án nhằm đem lại các sản phẩm giáo dục có chất lượng cao nhất cho học sinh và sinh viên, góp phần cải thiện nền giáo dục của Việt Nam. Vào thời điểm 05/2015, Tủ Sách Sputnik đã ra mắt bạn đọc 5 cuốn sách cho học sinh, và có kế hoach ra mắt hàng chục cuốn sách khác trong năm tiếp theo.

Các tác giả tin rằng Tủ Sách Sputnik gồm toàn những cuốn sách rất hay, được chọn lọc và dịch hoặc viết rất cẩn thận. Trong đó có những cuốn sách như *"Những cuộc phiêu lưu của người thích đếm"* nổi tiếng toàn thế giới, đã in ra hàng triệu bản, lần đầu xuất hiện ở Việt Nam. Có những cuốn sách nổi tiếng khác như *"Ba ngày ở nước Tí Hon"* trước đây đã từng được dịch ra tiếng Việt, nhưng bản dịch mới của Sputnik chính xác hơn, tránh được nhiều lỗi sai của bản dịch cũ. Bạn đọc sẽ không phí tiền khi mua chúng cho bản thân hay để tặng người thân.

Xin mời bạn đọc tìm hiểu kỹ hơn về Tủ Sách Sputnik ở phía cuối cuốn sách này. Các tác giả mong rằng bạn đọc sẽ nhiệt tình hưởn ứng Tủ Sách Sputnik, qua việc mua sách, quảng bá cho Tủ Sách Sputnik, v.v. Ủng hộ Tủ Sách Sputnik là một cách thiết thực để góp phần đem lại các sản phẩm giáo dục có chất lượng tốt hơn cho Việt Nam. Xin chân thành cảm ơn bạn đọc!

Hanoi–Toulouse, 05/2015

Lời giới thiệu

Xác suất và thống kê đóng vai trò rất quan trọng trong hầu hết mọi lĩnh vực của thế giới hiện đại, từ khoa học, công nghệ, đến kinh tế, chính trị, đến sức khỏe, môi trường, v.v. Ngày nay, máy tính giúp cho việc tính toán các vấn đề xác suất thống kê ngày càng trở nên dễ dàng, một khi đã có các số liệu đúng đắn và mô hình hợp lý. Thế nhưng, bản thân máy tính không biết mô hình nào là hợp lý. Đấy là vấn đề của người sử dụng: cần phải hiểu được bản chất của các khái niệm và mô hình xác suất thống kê, thì mới có thể dùng được chúng.

Mục đích của quyển sách này chính là nhằm giúp bạn đọc *hiểu* đúng bản chất của những khái niệm và phương pháp cơ bản nhất của xác suất và thống kê, và qua đó có thể áp dụng được chúng, đi sâu tìm hiểu được phương pháp thích hợp cho những tình huống cụ thể. Một số điểm mà các tác giả cố gắng đưa vào trong sách này là:

- Giải thích bản chất các khái niệm một cách trực giác, dễ hiểu nhất trong chừng mực có thể, đồng thời đảm bảo độ chặt chẽ nhất định về mặt toán học.

- Cho nhiều ví dụ và bài tập về những tình huống có thật, với số

liệu có thật, nhằm giúp bạn đọc cảm nhận được các ứng dụng thực tế của xác suất và thống kê.

Quyển sách này có 5 chương cộng thêm phần phụ lục. Chương 1 gồm một số khái niệm cơ sở của lý thuyết xác suất. Chương này không đòi hỏi kiến thức đặc biệt gì về toán, và học sinh phổ thông cũng có thể đọc và hiểu được phần lớn. Tuy nhiên, kiến thức của Chương 1 không hoàn toàn hiển nhiên, kể cả đối với những người đã học đại học. Trong quá trình soạn thảo, các tác giả có đem một số bài tập hơi khó của Chương 1 đố các học sinh đại học và cao học ngành toán, và phần lớn họ làm sai! Các bài tập đó không phải là khó về mặt toán học (để giải chúng chỉ cần làm vài phép tính số học đơn giản), mà là khó vì chúng chứa đựng những sự tế nhị về bản chất của xác suất. Hy vọng rằng, bạn đọc sẽ thấy được những sự tế nhị đó, và tránh được các sai lầm mà nhiều người khác hay mắc phải.

Từ Chương 2 đến Chương 4 của quyển sách là lý thuyết xác suất của các biến ngẫu nhiên. Chương 2 là về các biến ngẫu nhiên nhận giá trị thực. Chương 3 là về các bộ nhiều biến ngẫu nhiên, hay còn gọi là các vector ngẫu nhiên. Chương 4 là về các định lý giới hạn, trong đó có định lý giới hạn trung tâm, được coi là định lý quan trọng nhất của lý thuyết xác suất và là hòn đá tảng của thống kê toán học. Chương 5 của quyển sách là giới thiệu về thống kê. Bạn đọc sẽ tìm thấy trong chương này những vấn đề có thể giải quyết bằng thống kê như ước lượng, kiểm định, dự báo, những nguyên tắc cơ bản nhất của thống kê, và một số phương pháp thông kê nay đã trở thành kinh điển. Phụ lục A chứa lời giải của nhiều bài tập trong 4 chương đầu tiên của quyển sách.

Để hiểu tốt các vấn đề được bàn tới trong Chương 2 và các chương tiếp theo, bạn đọc cần có một số kiến thức chuẩn bị về giải tích toán học, như phép tính vi tích phân và khai triển Taylor-Lagrange, cộng với một ít kiến thức về đại số tuyến tính. Nếu có thêm một ít kiến thức về tôpô và giải tích hàm thì càng tốt. Trong sách có đưa ra định nghĩa và tính chất của một số khái niệm toán học cần dùng, ví dụ như tích phân Lebesgue trên không gian xác suất, biến đổi Fourier, hội tụ yếu, v.v.

Quyển sách này có thể dùng làm sách giáo khoa hay sách tham khảo cho môn xác suất thống kê ở bậc đại học hoặc cao học nhiều ngành khác nhau. Sinh viên các ngành không phải toán có thể bỏ qua các phần chứng minh các định lý tương đối phức tạp trong sách, mà chỉ cần hiểu đúng phát biểu của các định lý quan trọng nhất và cách áp dụng chúng. Các sinh viên ngành toán thì nên tìm hiểu cả cách chứng minh các định lý.

Do khuôn khổ của quyển sách có hạn, nên còn rất nhiều khái niệm quan trọng của xác suất và thống kê không xuất hiện trong sách, ví dụ như quá trình ngẫu nhiên, phương pháp bootstrap, hồi qui tuyến tính suy rộng, v.v.. Hy vọng rằng quyển sách này cung cấp được tương đối đầy đủ các kiến thức cơ sở, để bạn đọc có thể hiểu được các tài liệu chuyên sâu hơn về xác suất và thống kê khi cần thiết.

Để biên soạn quyển sách này, các tác giả có tham khảo nhiều sách báo liên quan đến xác suất thống kê, và có trích lại nhiều bài tập và ví dụ từ các tài liệu đó. Những sách mà các các tác giả tham khảo nhiều được liệt kê ở phần "Tài liệu tham khảo". Trong đó có những sách "nặng", có nhiều chứng minh chặt chẽ và khá nặng về toán,

ví dụ như quyển “Theory of probability and random processes” của Koralev và Sinai [5], và có những sách “nhẹ”, dễ đọc để có thể nắm được những ý tưởng chính, nhưng không có chứng minh, tiêu biểu như quyển “The cartoon guide to statistics” của Gonick và Smith [2].

Các hình minh họa trong quyển sách này chủ yếu được lấy từ internet. Chúng tôi tin rằng các hình đó thuộc phạm vi “public” và không bị hạn chế về mặt bản quyền, nhưng nếu do sơ suất mà chúng tôi sử dụng hình được bảo vệ bởi luật bản quyền mà chưa xin phép, thì chúng tôi xin thành thật xin lỗi trước.

Những bản thảo đầu tiên của quyển sách này có được một số đồng nghiệp, bạn bè và sinh viên đọc và góp ý sửa lỗi và trình bầy lại cho tốt lên. Các tác giả xin chân thành cảm ơn sự quan tâm và giúp đỡ của họ. Tất nhiên, mọi lỗi còn lại trong sách là thuộc về trách nhiệm của các tác giả. Đặc biệt, chúng tôi muốn cảm ơn các bạn Phan Thanh Hồng, Nguyễn Tuyết Mai, Nguyễn Thu Ngọc, Trần Quốc Tuấn và Lê Văn Tuấn, là các thành viên của Trung Tâm Toán Tài Chính và Công Nghiệp Hà Nội đã tích cực tham gia giúp chúng tôi soạn phần lời giải cho các bài tập.

Quyển sách này là một sản phẩm của Trung Tâm Toán Tài Chính và Công Nghiệp Hà Nội do các tác giả thành lập vào đầu năm 2009, được viết với mục đích trước hết là để phục vụ cho nhu cầu của bản thân Trung Tâm. Các tác giả hy vọng rằng, quyển sách này sẽ có ích, không chỉ cho Trung Tâm, mà còn cho một lượng rất lớn các độc giả khác đang hoặc sẽ quan tâm về xác suất và thống kê.

Hà Nội – Toulouse, 2010

Chương 1

Xác suất là gì

1.1 Xác suất là gì ?

Hầu như mọi người đều biết đến khái niệm xác suất. Tuy nhiên không phải ai cũng hiểu rõ những tính chất cơ bản của nó. Ví dụ như sự phụ thuộc vào thông tin của xác suất (mỗi khi có thêm thông tin mới thì xác suất thay đổi) hay bị bỏ qua. Và có những bài toán tính toán xác suất tưởng chừng như rất đơn giản, nhưng có hơn một nửa số người đã từng học xác suất làm sai khi được hỏi, kể cả các thạc sĩ ngành toán. Bởi vậy, trong chương này, chúng ta sẽ nhấn mạnh những sự tế nhị trong xác suất, đặc biệt là với xác suất có điều kiện, mà bạn đọc cần biết đến, để tránh được những lỗi cơ bản hay gặp nhất.

Trước khi đi vào lý thuyết, có một câu đố liên quan đến xác suất sau đây dành cho bạn đọc. Giả sử có một trò chơi trên TV như sau:

có 3 cánh cửa, đằng sau 1 trong 3 cánh cửa đó là 1 món quà lớn, còn sau 2 cửa còn lại không có gì. Người chơi được chọn 1 trong 3 cánh cửa, nếu chọn đúng cửa có quà thì được nhận quà. Sau khi người chơi đã chọn 1 cửa, người hướng dẫn chương trình mở một trong hai cửa còn lại ra, nhưng sẽ chỉ mở cửa không có quà. Sau đó người chơi được quyền chọn, hoặc là giữ cái cửa mình chọn ban đầu, hoặc là đổi lấy cái cửa chưa được mở còn lại. Theo bạn thì người chơi nên chọn phương án nào? Vì sao ? Hãy thử nghĩ về nó một chút trước khi tiếp tục đọc.

1.1.1 Xác suất của một sự kiện

Xác suất của một sự kiện (hay biến cố, tình huống giả định) là khả năng xảy ra sự kiện (hay biến cố, tình huống giả định) đó, được đánh giá dưới dạng một số thực nằm giữa 0 và 1.

Khi một sự kiện không thể xảy ra thì xác suất của nó bằng 0. Ví dụ như xác suất của sự kiện "có người sống trên sao Thổ" bằng 0.

Khi một sự kiện chắc chắn đã hoặc sẽ xảy ra thì xác suất của nó bằng 1 (hay còn viết là 100%). Ví dụ như sự kiện "tôi được sinh ra từ trong bụng mẹ" có xác suất bằng 1.

Khi một sự kiện có thể xảy ra và cũng có thể không xảy ra, và chúng ta không biết nó có chắn chắn xảy ra hay không, thì chúng ta có thể coi xác suất của nó lớn hơn 0 và nhỏ hơn 1. Sự kiện nào được coi là càng dễ xảy ra thì có xác suất càng lớn (càng gần 1), và ngược lại nếu càng khó xảy ra thì xác suất càng nhỏ (càng gần 0). Ví dụ tôi mua một vé xổ số. Tôi không biết nó sẽ trúng giải hay không, có thể

có mà cũng có thể không. Nếu như cứ 100 vé xổ số chỉ có 1 vé trúng giải, thì tôi sẽ coi xác suất trúng giải của vé của tôi là 1%. Con số 1% ở đây chính là tần suất, hay tỷ lệ trúng giải của các vé xổ số: nó bằng số các vé trúng giải chia cho tổng số các vé.

Không những chỉ các sự kiện trong tương lai, mà cả các sự kiện trong quá khứ, mà chúng ta thiếu thông tin để có thể biết chắc là chúng đã thực sự xảy ra hay không, thì chúng ta vẫn có thể gán cho các sự kiện đó một xác suất nào đó, ứng với độ tin tưởng của chúng ta về việc sự kiện đó đã thực sự xảy ra hay không. Ví dụ như, nữ hoàng Cleopatra của Ai Cập có tự tử bằng cách để cho rắn độc cắn không ? Đấy là một giả thuyết, mà theo các nhà sử học thì có nhiều khả năng xảy ra, nhưng không chắc chắn.

1.1.2 Ba tiên đề về sự nhất quán của xác suất

Tiên đề 1. Như đã viết phía trên, nếu A là một sự kiện (giả định) và ký hiệu $P(A)$ là **xác suất của** A thì

$$0 \leq P(A) \leq 1 \tag{1.1}$$

Tiên đề 2. Nếu A là một sự kiện, và ký hiệu $\overline{A}$ là sự kiện *phủ định của* A thì

$$P(A) + P(\overline{A}) = 1 \tag{1.2}$$

Ý nghĩa triết học của tiên đề 2 tương đối hiển nhiên: Trong hai sự kiện "A" và "phủ định của A" có 1 và chỉ 1 sự kiện xảy ra. Nếu "A" càng có nhiều khả năng xả ra thì "phủ định của A" càng có ít khả năng xảy ra, và ngược lại.

Ví dụ 1.1. Một học sinh đi thi vào một trường đại học. Nếu xác suất thi đỗ là 80% thì xác suất thi trượt là 20% (= 100% - 80%), chứ không thể là 30%, vì nếu xác suất thi đỗ là 80% và xác suất thi trượt là 30% thì không nhất quán.

Ví dụ 1.2. Tôi tung một đồng tiền, khi nó rơi xuống thì có thể hiện mặt sấp hoặc mặt ngửa. Tổng xác suất của hai sự kiện "mặt sấp" và "mặt ngửa" bằng 1. Nếu tôi không có lý do đặc biệt gì để nghĩ rằng mặt nào dễ hiện lên hơn mặt nào, thì tôi coi rằng hai mặt có xác suất hiện lên bằng nhau. Khi đó sự kiện "mặt ngửa" có xác suất bằng sự kiện "mặt sấp" và bằng $1/2$.

Tiên đề 3. Với hai sự kiện A và B, ta sẽ ký hiệu sự kiện "cả A và B đều xảy ra" bằng $A \cap B$ và sự kiện "ít nhất một trong hai sự kiện A hoặc B xảy ra" bằng $A \cup B$. Khi đó nếu hai sự kiện A và B không thể cùng xảy ra, thì xác suất của sự kiện "xảy ra A hoặc B" bằng tổng các xác suất của A và của B:

$$P(A \cap B) = 0 \Rightarrow P(A \cup B) = P(A) + P(B) \tag{1.3}$$

Ví dụ 1.3. Một học sinh được cho điểm một bài kiểm tra. Có thể được 7 điểm, có thể được 8 điểm, hoặc có thể được điểm khác, nhưng không thể vừa được 7 điểm vừa được 8 điểm. Bởi vậy $P((7d)\cup(8d)) = P(7d) + P(8d)$

Tiên đề 3 có thể phát biểu một cách tổng quát hơn như sau:

Tiên đề 3'. Nếu X và Y là hai sự kiện bất kỳ thì

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{1.4}$$

Bài tập 1.1. Chứng minh rằng tiên đề 3 tương đương với tiên đề 3'.

1.1.3 Xác suất phụ thuộc vào những gì ?

Xác suất của một sự kiện không nhất thiết phải là một hằng số, mà nó có thể thay đổi, phụ thuộc vào nhiều yếu tố. (Từ *sự kiện* ở đây hiểu theo nghĩa thông thường, chứ không phải theo nghĩa "một tập hợp trong một không gian xác suất với 1 độ đo xác suất đã cố định" trong mô hình toán học)

Xác suất thay đổi theo thời gian. Ví dụ, ông Obama được bầu làm tổng thống Mỹ vào tháng 11/2008. Từ trước lúc bầu cử mấy tháng, có sự cạnh tranh ác liệt giữa ông ta và đối thủ chính của ông ta là ông McCain, và một người quan sát bên ngoài có thể nhận định là hai ông có khả năng được bầu cử ngang nhau (tức là xác suất được bầu của mỗi ông quãng 50%). Nhưng khi kết quả bầu cử được công bố trọn vẹn, thì xác suất được bầu của Obama chuyển thành 100% (tức là ông ta đã chắc chắn được bầu). Trước đó 1 năm, ông Obama là một người chưa được nhiều người biết đến và còn phải tranh cử với bà Clinton và các ứng cử viên khác trong Đảng của mình, và khi đó, đối với quan sát viên bên ngoài, xác suất được bầu làm tổng thống của Obama không phải 100%, cũng không phải 50%, mà nhỏ hơn thế nhiều.

Xác suất phụ thuộc vào thông tin. Lấy bài toán đố về trò chơi trên TV viết phía trên làm ví dụ. Gọi tên cửa mà người chơi chọn lúc đầu là A, cửa không có quà mà người hướng dẫn chương trình mở ra là B, và cửa còn lại là C. Vào thời điểm ban đầu, không có thông tin gì về cửa nào phía sau có quà, thông tin duy nhất là 1 trong 3 cửa có quà. Không có cơ sở gì để cho rằng cửa nào có nhiều khả năng có quà

hơn cửa nào, bởi vậy vào thời điểm ban đầu ta coi $P(A) = P(B) = P(C) = 1/3$. Nhưng sau khi cửa B được mở ra, thì ta có thêm một thông tin mới, là cửa B không có quà. Như vậy thông tin mới này làm thay đổi xác suất của B: bây giờ ta có $P(B) = 0$. Không chỉ xác suất của B thay đổi, mà tổng xác suất của A và C bây giờ cũng thay đổi: $P(A) + P(C) = 1$ thay vì bằng 2/3 như trước. Như vậy ít ra một trong hai số $P(A)$ hoặc $P(C)$ thay đổi, hoặc là cả hai. Xác suất $P(A)$ có thay đổi vì thông tin mới này không ? Câu trả lời là không (Giải thích vì sao không ?). Chỉ có $P(C)$ là thay đổi: sau khi người hướng dẫn chương trình mở cửa B, thì ta có $P(A) = 1/3$ và $P(C) = 2/3$. Như vậy người chơi nên đổi cửa A lấy cửa C thì dễ thắng hơn. Để thấy rõ hơn việc cánh cửa còn lại có nhiều khả năng có quà hơn là cánh cửa mà người chơi chọn ban đầu, thay vì chỉ có 3 cửa, ta hãy hình dung có 100 cửa. Sau khi bạn chọn 1 cửa, người dẫn chương trình mở 98 cửa không có quà trong số 99 cửa còn lại, chỉ để lại 1 cửa thôi. Khi đó, nếu được đổi, bạn sẽ giữ nguyên cửa của mình, hay là đổi lấy cái cửa còn lại kia ?

Xác suất phụ thuộc vào điều kiện. Chúng ta sẽ bàn về xác suất có điều kiện và công thức tính xác suất có điều kiện ở một phần sau. Điều đáng chú ý ở đây là, mọi xác suất đều có thể coi là xác suất có điều kiện, và đều phụ thuộc vào những điều kiện nào đó, có thể được nói ra hoặc không nói ra (điều kiện hiểu ngầm). Ví dụ, khi chúng ta nói "khi tung cái xúc sắc S, xác suất để hiện lên mặt có 3 chấm là 1/6", chúng ta hiểu ngầm S là một cái xúc sắc đều đặn, các mặt đều có khả năng xuất hiện như nhau. Nhưng nếu S là một cái xúc sắc méo mó, nhẹ bên này nặng bên nọ (điều kiện khác đi), thì hoàn toàn

có thể là xác suất để khi tung hiện lên mặt có 3 chấm sẽ khác 1/6. Một ví dụ khác là xác suất xảy ra tai nạn khi lái ô tô: khi người lái xe khoe mạnh tỉnh táo, thì xác suất xảy ra tai nạn thấp, còn khi vẫn người lái đó bị say rượu hoặc buồn ngủ gật, thì xác suất xảy ra tai nạn cao hơn, v.v. Khi chúng ta biết thêm một điều kiện mới, tức là có thêm một thông tin mới, bởi vậy sự phụ thuộc vào điều kiện của xác suất cũng có thể coi là sự phụ thuộc vào thông tin.

Xác suất phụ thuộc vào người quan sát, hay là tính chủ quan của xác suất. Cùng là một sự kiện, nhưng hai người quan sát khác nhau có thể tính ra hai kết quả xác suất khác nhau, và cả hai đều "có lý", bởi vì họ dựa trên những thông tin và phân tích khác nhau. Ví dụ như, có chuyên gia tài chính đánh giá rằng cổ phiếu của hãng Vinamilk có nhiều khả năng đi lên trong thời gian tới, trong khi lại có chuyên gia tài chính khác đánh giá rằng cổ phiếu của hãng đó có nhiều khả năng đi xuống ít khả năng đi lên trong thời gian tới. Quay lại trò chơi truyền hình: với người chơi thì $P(A) = 1/3$, nhưng đối với người dẫn chương trình thì $P(A)$ không phải là 1/3, mà là 0 hoặc 1, vì người đó biết ở đằng sau cửa A có quà hay không.

1.1.4 Tính xác suất bằng thống kê

Đối với những hiện tượng xảy ra nhiều lần, thì người ta có thể dùng thống kê để tính xác suất của sự kiện xảy ra hiện tượng đó. Công thức sẽ là

$$P(A) = \frac{N(A)}{N(total)} \tag{1.5}$$

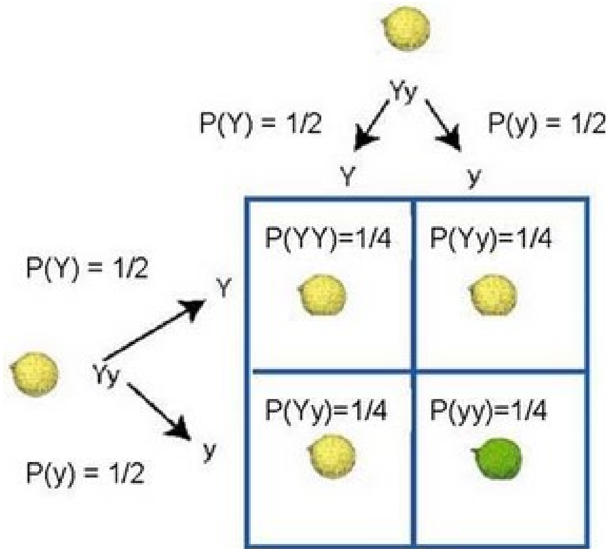
Ở đây $N(total)$ là tổng số các trường hợp được khảo sát, và $N(A)$ là số các trường hợp được khảo sát thỏa mãn điều kiện *xảy ra* A.

Cơ sở toán học cho việc dùng thống kê để tính xác suất, là luật số lớn và các định lý giới hạn, mà chúng ta sẽ tìm hiểu ở phía sau trong sách này.

Ví dụ 1.4. Có một số số liệu sau đây về tai tạn ô tô và máy bay. Trong những năm 1989-1999, trên toàn thế giới, trung bình mỗi năm có khoảng 18 triệu chuyến bay, 24 tai nạn máy bay chết người, và 750 người chết trong tai nạn máy bay. Cũng trong khoảng thời gian đó, ở nước Pháp, trung bình mỗi năm có khoảng 8000 người chết vì tai nạn ô tô, trên tổng số 60 triệu dân. Từ các số liệu này, chúng ta có thể tính: Xác suất để một người ở Pháp bị chết vì tai nạn ô tô trong một năm là 8000/60000000 = 0,0133%. Xác suất để đi một chuyến bay gặp tai nạn chết người là 24/18000000 = 0,000133%, chỉ bằng 1/100 xác suất bị chết vì tai nạn ô tô trong 1 năm. Nếu một người một năm bay 20 chuyến, thì xác suất bị chết vì tai nạn máy bay trong năm bằng quãng $20 \times 0,000133\% = 0,00266\%$, tức là chỉ bằng 1/5 xác suất bị chết vì tai nạn ô tô trong năm.

Ví dụ 1.5. Ông Gregor Mendel (1822-1884) là một tu sĩ người Áo (Austria) thích nghiên cứu sinh vật. Ông ta trồng nhiều giống đậu khác nhau trong vườn của tu viện, và ghi chép tỉ mẩn về các tính chất di truyền và lai giống của chúng. Năm 1866 Mendel công bố một bài báo về các hiện tượng mà ông ta qua sát được, và lý thuyết của ông ta để giải thích các hiện tượng. Một trong những quan sát trong đó là về màu sắc: Khi lai đậu hạt vàng với đậu hạt xanh (thế hệ thứ nhất) thì các cây lai (thế hệ thứ hai) đều ra đậu hạt vàng,

nhưng tiếp tục lai các cây đậu hạt vàng thế hệ thứ hai này với nhau, thì đến thế hệ thứ ba xác suất ra đậu hạt xanh là 1/4. Con số 1/4



Hình 1.1: Lý thuyết di truyền của Mendel và xác suất trong lai giống đậu

là do Mendel thống kê thấy tỷ lệ đậu hạt xanh ở thế hệ thứ ba gần bằng 1/4. Từ đó Mendel xây dựng lý thuyết di truyền để giải thích hiện tượng này: màu của đậu được xác định bởi 1 gen, và gen gồm có hai phần. Thế hệ đầu tiên, cây đậu hạt vàng có gen thuần chủng "YY" còn hạt xanh có gen "yy" (tên gọi "Y" và "y" ở đây là tùy tiện). Khi lai nhau, thì một nửa gen của cây này ghép với một nửa gen của cây kia để tạo thành gen của cây con. Các cây thế hệ thứ hai đều có gen "Yy", và màu hạt của gen "Yy" cũng là vàng. Đến thế hệ thứ ba,

khi lai "Yy" với "Yy" thì có 4 khả năng xảy ra : "YY", "Yy", "yY" và "yy". ("Yy" và "yY" là giống nhau về gen, nhưng viết như vậy là để phân biệt là phần "Y" đến từ cây thứ nhất hay cây thứ hai trong 2 cây lai với nhau). Về lý thuyết, có thể coi 4 khả năng trên là có xác suất xảy ra bằng nhau. Bởi vậy xác suất để cây thế hệ thứ ba có gen "yy" (hạt màu xanh) là 1/4. Trong rất nhiều năm sau khi công bố, công trình của Mendel không được các nhà khoa học khác quan tâm đến, nhưng ngày nay Mendel được coi là cha tổ của di truyền học.

1.2 Mô hình toán học của xác suất

1.2.1 Không gian xác suất

Không gian xác suất là một khái niệm toán học nhằm trừu tượng hóa 3 tiên đề phía trên về sự nhất quán của xác suất.

Định nghĩa 1.1. *Một* **không gian xác suất** *là một tập hợp Ω, cùng với:*

1) Một họ $\mathcal{S}$ các tập con của Ω, thỏa mãn các tính chất sau: $\Omega \in \mathcal{S}$, và nếu $A, B \in \mathcal{S}$ thì $A \cup B \in \mathcal{S}$, $A \cap B \in \mathcal{S}$ và $\overline{A} := \Omega \setminus A \in \mathcal{S}$. Một họ như vậy được gọi là một **đại số** *các tập con của Ω. Trong trường hợp Ω là một tập có vô hạn các phần tử, thì chúng ta sẽ đòi hỏi thêm điều kiện sau: Nếu $A_i, i = 1, 2, 3, \ldots$ là một dãy vô hạn các phần tử của $\mathcal{S}$, thì hợp $\bigcup_{i=1}^{\infty} A_i$ cũng thuộc họ $\mathcal{S}$. Với thêm điều kiện này, $\mathcal{S}$ được gọi là một* **sigma-đại số**. *Các phần tử của $\mathcal{S}$ được gọi là là tập hợp con* **đo được** *của không gian xác suất.*

2) Một hàm số thực $P : \mathcal{S} \to \mathbb{R}$ trên $\mathcal{S}$, được gọi là **phân bố xác**

suất *hay* **độ đo xác suất** *trên Ω, thỏa mãn các tính chất sau:*
i) Với mọi $A \in \mathcal{S}$, ta có

$$0 \leq P(A) \leq 1. \tag{1.6}$$

ii)

$$P(\emptyset) = 0,\ P(\Omega) = 1. \tag{1.7}$$

iii) Nếu $A \cap B = \emptyset$ thì

$$P(A \cup B) = P(A) + P(B). \tag{1.8}$$

Tổng quát hơn, nếu $A_i, i = 1, 2, 3, \ldots$ là một dãy các tập hợp con đo được không giao nhau thì

$$P(\bigcup_i A_i) = \sum_i P(A_i). \tag{1.9}$$

Ghi chú 1.1. 1) Không gian xác suất Ω còn được gọi là **không gian mẫu** (sample space), và nó là mô hình toán học trừu tượng cho vấn đề tính toán xác suất đang được quan tâm. Mỗi phần tử của Ω có thể được gọi là một **sự kiện thành phần** (elementary event). Nếu A là một phần tử của Ω thì ta cũng có thể viết $P(A)$ và hiểu là $P(\{A\})$, trong đó $\{A\}$ là tập con của Ω chứa duy nhất một phần tử A. Mỗi sự kiện là một tập con của Ω, và có thể gồm nhiều (thậm chí vô hạn) sự kiện thành phần. Không nhất thiết tập con nào của Ω cũng đo được (tức là nằm trong họ $\mathcal{S}$), và chúng ta sẽ chỉ quan tâm đến những tập con đo được.

2) Trong toán học, một đại số là một tập hợp với các phép tính cộng, trừ, và phép nhân (không nhất thiết phải có phép chia). Các tính chất của họ $\mathcal{S}$ trong định nghĩa không gian xác suất khiến nó là một đại



Hình 1.2: A. N. Kolmogorov

số theo nghĩa như vậy: Phần tử 0 trong $\mathcal{S}$ là tập rỗng, phần tử đơn vị trong $\mathcal{S}$ là tập Ω, phép nhân trong $\mathcal{S}$ là phép giao: $A \times B := A \cap B$, và phép cộng trong $\mathcal{S}$ là phép $A+B := (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$. Đại số này có số đặc trưng bằng 2, tức là $2A = A + A = 0$ với mọi A (và bởi vậy phép cộng và phép trừ chẳng qua là một). Chúng ta muốn $\mathcal{S}$ là một đại số chính là để cho việc làm các phép tính số học với xác suất được thuận tiện.

3) Đẳng thức (1.9) được gọi là **tính chất sigma** của xác suất. Trong toán, chữ cái hy lạp sigma thường dùng để ký hiệu tổng, với hữu hạn hay vô hạn các thành phần. Tính chất sigma là *tính chất cộng tính vô*

hạn: khi có một dãy vô hạn các tập con không giao nhau, xác suất của hợp của chúng cũng bằng tổng vô hạn của các xác suất của các tập con. Tính chất sigma chính là tính chất cho phép chúng ta *lấy giới hạn* trong việc tính toán xác suất. Chẳng hạn như, nếu $A_1 \subset A_2 \subset \dots$ là một dãy tăng các tập con của Ω, và $A = \lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$, thì ta có thể viết $P(A) = \lim_{n\to\infty} P(A_n)$, bởi vì

$$P(A) = P(A_1 \cup \bigcup_{n=1}^{\infty}(A_{n+1} \setminus A_n)) = P(A_1) + \sum_{n=1}^{\infty} P(A_{n+1} \setminus A_n)$$
$$= P(A_1) + \lim_{n\to\infty} \sum_{k=1}^{n} P(A_{k+1} \setminus A_k) = P(A_1) + \lim_{n\to\infty}(P(A_{n+1}) - P(A_1)) \quad (1.10)$$

Phép toán *lấy giới hạn* là phép toán cơ bản nhất của giải tích toán học, và mọi phép toán giải tích khác như đạo hàm, tích phân, v.v. đều có thể được định nghĩa qua phép lấy giới hạn. Bởi vậy, tính chất *sigma* chính là tính chất cho phép chúng ta sử dụng giải tích toán học trong việc nghiên cứu xác suất. Các nhà toán học cổ điển trong thế kỷ 18 và 19 đã dùng các phép tính vi tích phân trong xác suất, tức là đã dùng tính chất sigma. Về mặt trực giác, tính chất sigma là mở rộng hiển nhiên của tính chất cộng tính (1.8). Tuy nhiên, nói một cách chặt chẽ toán học, đẳng thức (1.9) không suy ra được từ đẳng thức (1.8), và phải được coi là một tiên đề trong xác suất. Tiên đề này được đưa ra bởi nhà toán học người Nga Andrei Nikolaievitch Kolmogorov (1903-1987), người xây dựng nền tảng cho lý thuyết xác suất hiện đại.

Bài tập 1.2. Chứng minh rằng, với 3 tập con A, B, C (đo được) bất

kỳ trong một không gian xác suất, ta có:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) \\ - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C).$$

1.2.2 Phân bố xác suất Bernoulli



Hình 1.3: Bia mộ của "mathematicus incomparabilis" J. Bernoulli ở Basel

Không gian xác suất đơn giản nhất mà không tầm thường là

không gian sinh bởi đúng 1 sự kiện A và phủ định $\overline{A}$ của nó: $\Omega = \{A, \overline{A}\}$. Phân bố xác suất trên Ω trong trường hợp này được xác định bởi đúng một số $p = P(A)$. Phân bố này được gọi là **phân bố Bernoulli**, theo tên của Jacob Bernoulli (1654-1705), một nhà toán học người Thụy Sĩ.

Ví dụ 1.6. Một vận động viên bắn súng, nhằm vào đích bắn 1 phát súng. Có hai sự kiện đối lập nhau có thể xảy ra là $A =$ "bắn trúng" và $\overline{A} =$ "bắn trượt". Giả sử xác suất bắn trúng là 95%. Khi đó ta có không gian xác suất $\Omega = \{A, \overline{A}\}$ với phân bố xác suất Bernoulli với $p = P(A) = 95\%$. Xác suất của $\overline{A}$ (sự kiện "bắn trượt") bằng $1 - p = 1 - 95\% = 5\%$.

Ví dụ 1.7. (Cái kim của Buffon). Bá tước George-Louis Leclerc de Buffon (1707-1788) là một nhà khoa học tự nhiên lớn, nghiên cứu về thực vật, động vật, trái đất, lịch sử tự nhiên, v.v. Thời trẻ, ông ta đặc biệt thích toán học, và vào năm 1733 có trình lên Viện Hàm lâm Pháp một công trình nhan đề "Sur le jeu du franc-carreau" (về trò chơi franc-carreau, là một trò chơi cá cược thịnh hành thời đó: người ta tung 1 đồng tiền vào 1 ô vuông và cá cược nhau xem vị trí nó sẽ nằm chỗ nào). Trong công trình này, các phép toán vi tích phân được Buffon đưa vào lý thuyết xác suất. Buffon còn là người nghĩ ra phương pháp sau đây để tính số π: Lấy 1 tờ giấy to và 1 cái kim. Kẻ các đường thẳng song song trên tờ giấy, cách đều nhau một khoảng cách đúng bằng chiều dài của cái kim. Tung cái kim một cách ngẫu nhiên lên trên tờ giấy. Có hai khả năng xảy ra: 1) kim nằm đè lên 1 đường thẳng trong các đường được kẻ; 2) kim nằm lọt vào giữa hai đường thẳng. Buffon tính ra rằng, sự kiện "kim nằm đè lên 1 đường

thẳng" có xác suất bằng $1/\pi$. Như vậy hai sự kiện "nằm đè lên 1 đường thẳng" và "nằm lọt vào giữa hai đường thẳng" hợp thành một không gian xác suất Bernoulli với $p = 1/\pi$. Tung kim n lần, và gọi số lần kim nằm đè lên 1 đường thẳng trong số n lần tung là b_n. Khi đó, theo luật số lớn, b_n/n tiến tới $p = 1/\pi$ khi n tiến tới vô cùng. Bởi vậy để xấp xỉ tính số π, có thể làm như sau: tung kim thật nhiều lần, đếm số lần kim đè lên trên 1 đường thẳng, rồi lấy số lần tung chia cho số đó. Phương pháp tung kim của Buffon chính là tiền thân của phương pháp Monte-Carlo trong toán học.

1.2.3 Phân bố xác suất đều

Định nghĩa 1.2. *Phân bố xác suất P trên không gian xác suất hữu hạn với N phần tử $\Omega = \{A_1, \ldots, A_N\}$ được gọi là* **phân bố xác suất đều** *nếu như $P(A_1) = \ldots = P(A_N) = 1/N$.*

Tất nhiên, mỗi không gian xác suất với một số hữu hạn các phần tử chỉ có duy nhất một phân bố xác suất đều trên đó.

Ghi chú 1.2. Khái niệm phân bố đều không mở rộng được lên các không gian xác suất có số phần tử là vô hạn và đếm được, bởi vì 1 chia cho vô cùng bằng 0, nhưng mà tổng của một chuỗi vô hạn số 0 vẫn bằng 0 chứ không bằng 1.

Các phân bố xác suất đều là các phân bố quan trọng hay gặp trong thực tế. Lý do chính dẫn đến phân bố xác suất đều là *tính đối xứng, cân bằng*, hay *hoán vị được* của các sự kiện thành phần.

Ví dụ 1.8. Lấy một bộ bài tú lơ khơ mới có 52 quân, đặt nằm sấp. Khi đó xác suất để rút một con bài trong đó ra một cách tùy ý được con



Hình 1.4: Tượng của Buffon ở Jardin des Plantes, Paris

"2 Cơ" (hay bất kỳ "số" nào khác) bằng 1/52. Vì sao vậy? Vì các con bài khi đặt nằm sấp thì giống hệt nhau, không thể phân biệt được con nào với con nào, số nào cũng có thể được viết dưới bất kỳ con bài nào, và nếu chuyển chỗ 2 con bài trong bộ bài với nhau thì trông bộ bài vẫn hệt như cũ (đấy chính là tính "đối xứng", "hoán vị được"). Người quan sát không có thông tin gì để có thể nhận biết được số

nào dễ nằm ở phía dưới con bài nào hơn trong các con bài đăng nằm sấp, và khi đó thì phải coi rằng xác suất của các số là như nhau. Nếu như có những con bài "được đánh dấu" (chơi ăn gian), thì tất nhiên đối với người biết chuyện đánh dấu, không còn phân bố xác suất đều nữa.

Công thức để tính xác suất của một sự kiện trong một phân bố xác suất đều rất đơn giản: Nếu như không gian xác suất Ω với phân bố xác suất đều có N phần tử, và sự kiện được biểu diễn bằng một tập con A của Ω với k phần tử, thì xác suất của A bằng k/N:

$$P(A) = \frac{\#A}{\#\Omega} = \frac{k}{N} \tag{1.11}$$

Ví dụ 1.9. Giả sử một gia đình có 3 con. Khi đó xác suất để gia đình đó có 2 con trai 1 con gái là bao nhiêu. Chúng ta có thể lập mô hình xác suất với 4 sự kiện thành phần: 3 trai, 2 trai 1 gái, 1 trai 2 gái, 3 gái. Thế nhưng 4 sự kiện thành phần đó không "cân bằng" với nhau, và bởi vậy không kết luận được rằng xác suất của "2 trai 1 gái" là 1/4. Để có không gian xác suất với phân bố đều, ta có thể lập mô hình xác suất với 8 sự kiện thành phần như sau:

$$\Omega = \{TTT, TTG, TGT, TGG, GTT, GTG, GGT, GGG\}.$$

(Chẳng hạn, GGT có nghĩa là con thứ nhất là con gái, con thứ hai là con gái, con thứ ba là con trai). Sự kiện "2 trai mội gái" là hợp của 3 sự kiện thành phần trong mô hình xác suất này: TTG, TGT, GTT. Như vậy xác suất của nó bằng 3/8.

Bài tập 1.3. Có một nhóm n bạn, trong đó có hai bạn Vôva và Lily. Xếp các bạn trong nhóm thành một hàng dọc một cách ngẫu nhiên. Hỏi xác suất để Vôva ở vị trí ngay sau Lily trong hàng là bao nhiêu?

Bài tập 1.4. Một nhóm có 5 người, với 5 tên khác nhau. Mỗi người viết tên của một người khác trong nhóm một cách ngẫu nhiên vào giấy. Tính xác suất để có 2 người trong nhóm viết tên của nhau.

Bài tập 1.5. Giả sử trong một giải bóng đá đấu loại trực tiếp có 8 đội A, B, C, D, E, F, G, H tham gia: vòng 1 có 4 trận, vòng 2 có 2 trận, vòng 3 (vòng cuối cùng) có 1 trận. Giá sử xác suất để mỗi đội thắng mỗi trận đều là 1/2, và các đội bắt thăm để xem đội nào đấu với đội nào ở vòng đầu, các vòng sau thì được xếp theo kết quả vòng trước. Tính xác suất để đội A có đấu với đội B trong giải.

1.2.4 Mô hình xác suất với vô hạn các sự kiện

Mọi vấn đề xuất phát từ thực tế đều chỉ có một số hữu hạn các sự kiện thành phần. Nhưng khi mà số sự kiện thành phần đó lớn, thì người ta có thể dùng các mô hình toán học với vô hạn phần tử để biểu diễn, cho dễ hình dung và tiện tính toán.

Ví dụ 1.10. Nếu ta quan tâm đến lượng khách hàng trong một ngày của một siêu thị, thì có thể dùng tập hợp các số nguyên không âm $\mathbb{Z}_+$ làm không gian xác suất: mỗi số $n \in \mathbb{Z}_+$ ứng với một sự kiện "số khách trong ngày là n". Vấn đề tiếp theo là chọn phân bố xác suất nào trên $\mathbb{Z}_+$ cho hợp lý (phản ánh khá chính xác thực tế xảy ra, đồng thời lại tiện cho việc tính toán)? Ví dụ người ta có thể dùng phân bố xác suất sau trên $\mathbb{Z}_+$, gọi là phân bố Poisson (đọc là Poa-Sông): $P(n) = e^{-\lambda}\dfrac{\lambda^n}{n!}$ với mọi $n \in \mathbb{Z}_+$. (Chú ý rằng $\sum_n P(n) = \sum_n e^{-\lambda}\dfrac{\lambda^n}{n!} = e^{-\lambda}\sum_n \dfrac{\lambda^n}{n!} = e^{-\lambda}e^{\lambda} = 1$, như vậy các tiên đề về xác suất được thỏa

mãn). Phân bố Poisson ứng với hai giả thuyết: lượng khách hàng trung bình trong một ngày là λ, và các khách hàng đi đến siêu thị một cách ngẫu nhiên và độc lập với nhau. Chúng ta sẽ tìm hiểu kỹ hơn về phân bố Poisson trong những phần sau.

Ví dụ 1.11. Ta biết rằng có một xe ô tô X đang đậu ở trên một khúc phố Z, và ta quan tâm đến vị trí của X trên phố đó. Ta có thể mô hình X bằng 1 điểm, Z bằng một đoạn thẳng và lấy đoạn thẳng đó làm không gian xác suất: $\Omega = [a, b]$, $a, b \in \mathbb{R}$, $a < b$. (Mô hình xác suất liên tục này có số phần tử là continuum, không đếm được). Sự kiện "ô tô đỗ ở chỗ nào đó trên khúc phố" chuyển thành sự kiện "điểm x nằm trong một đoạn thẳng con nào đó trên đoạn thẳng $\Omega = [a, b]$". Ta có thể chọn phân bố xác suất đều trên $\Omega = [a, b]$ theo nghĩa sau: xác suất của mỗi đoạn thẳng con trên Ω tỷ lệ thuận với độ dài của đoạn thẳng con đó, và bằng chiều dài của đoạn thẳng con đó chia cho chiều dài của Ω: $P([c, d]) = (d - c)/(b - a)$.

1.2.5 Ánh xạ giữa các không gian xác suất

Cùng một vấn đề tính toán xác suất, ta có thể lập nhiều mô hình không gian xác suất khác nhau. Ví dụ, mô hình xác suất đơn giản nhất cho sự kiện "bị ốm" sẽ là mô hình Bernoulli $\Omega_1 = \{S, H\}$ với 2 sự kiện S = "bị ốm" (sick) và H = "không bị ốm" (healthy). Như ta cũng có thể chia nhỏ sự kiện bị ốm ra thành rất nhiều sự kiện con, ví dụ như "ốm bệnh A", "ốm bệnh B", "ốm cả bệnh A lẫn bệnh B", v.v. và sự kiện "không bị ốm" cũng có thể chia thành nhiều sự kiện con, ví dụ như "rất khỏe", "không ốm nhưng mà yếu", v.v. Khi chia nhỏ như vậy, ta được mô hình xác suất với một không gian xác

suất $\Omega_2 = \{S_1, S_2, \ldots, H_1, H_2, \ldots\}$ với nhiều phần tử hơn. Hai không gian đó liên quan với nhau bởi một ánh xạ $\phi : \Omega_2 \to \Omega_1$, $\phi(S_i) = S, \phi(H_i) = H$. Tất nhiên, khi ta chia nhỏ sự kiện S ra thành nhiều sự kiện (không giao nhau) $S_1, S_2, \ldots$, thì không phải vì thế mà xác suất của nó thay đổi. Nói cách khác, ta phải có

$$P(S) = P(\phi^{-1}(S)) = P(\cup_i S_i) = \sum_i P(S_i) \tag{1.12}$$

Tính chất trên là tính chất bảo toàn xác suất của ánh xạ ϕ. Nói một cách tổng quát, ta có định nghĩa sau:

Định nghĩa 1.3. *Một ánh xạ $\phi : (\Omega_1, P_1) \to (\Omega_2, P_2)$ từ một không gian xác suất (Ω_1, P_1) vào một không gian xác suất (Ω_2, P_2) được gọi là một* **ánh xạ bảo toàn xác suất** *nếu nó bảo toàn độ đo xác suất, có nghĩa là với mọi tập con $B \subset \Omega_2$ đo được, ta có*

$$P_1(\phi^{-1}(B)) = P_2(B) \tag{1.13}$$

Nếu hơn nữa, ϕ là một song ánh modulo những tập có xác suất bằng 0, có nghĩa là tồn tại các tập con $A \in \Omega_1$, $B \in \Omega_2$ sao cho $P_1(A) = P_2(B) = 0$ và $\phi : \Omega_1 \setminus A \to \Omega_2 \setminus B$ là song ánh bảo toàn xác suất), thì ϕ được gọi là một **đẳng cấu xác suất** *, và ta nói rằng (Ω_1, P_1) đẳng cấu xác suất với (Ω_2, P_2).*

Ví dụ 1.12. Đặt 4 bạn Al, Ben, Cam, Don ngồi vào 4 ghế A, B, C, D một cách hoàn toàn ngẫu nhiên. Tính xác suất để Al được đặt ngồi vào ghế A. Có 4 ghế, và xác suất để Al ngồi vào mỗi nghế trong 4 ghế đó coi là bằng nhau (vì không có cớ gì để coi là khác nhau), bởi vậy xác suất để Al ngồi vào ghế A là 1/4. Nhưng cũng có thể lý luận tỷ

mẩn hơn như sau: có tổng cộng 4! = 24 cách đặt 4 bạn ngồi vào 4 ghế, trong đó có 3! = 6 cách có Al ngồi vào ghế A. Bởi vậy xác suất để Al ngồi vào ghế A là 6/24 = 4. Hai cách giải cho cùng một đáp số, nhưng sử dụng hai không gian xác suất khác nhau: không gian thứ nhất có 4 phần tử, còn không gian thứ hai có 24 phần tử. Có một phép chiếu tự nhiên bảo toàn xác suất từ không gian thứ hai lên không gian thứ nhất.

Định lý 1.1. *Nếu (Ω_1, P_1) là một không gian xác suất, và $\phi : \Omega_1 \to \Omega_2$ là một ánh xạ tùy ý, thì tồn tại một độ đo xác suất P_2 trên Ω_2, sao cho ánh xạ $\phi : (\Omega_1, P_1) \to (\Omega_2, P_2)$ là ánh xạ bảo toàn xác suất.*

Chứng minh. Có thể xây dựng P_2 theo công thức sau: với mỗi tập con $B \subset \Omega_2$, nếu tồn tại $P_1(\phi^{-1}(B))$ thì ta đặt

$$P_2(B) := P_1(\phi^{-1}(B)) \tag{1.14}$$

Độ đo xác suất P_2 định nghĩa theo công thức trên được gọi là **push-forward** của P_1 qua ánh xạ ϕ, hay còn gọi là **phân bố xác suất cảm sinh** từ P_1 qua ánh xạ ϕ. □

Bài tập 1.6. Chứng minh rằng quan hệ đẳng cấu xác suất giữa các không gian xác suất là một quan hệ tương đương.

1.2.6 Tích của các không gian xác suất

Nếu M và N là hai tập hợp, thì tích của chúng (hay còn gọi là tích trực tiếp, hay tích Descartes), ký hiệu là $M \times N$, là tập hợp các cặp phần tử (x, y), $x \in M, y \in N$. Trong trường hợp $M = (\Omega_1, P_1)$ và $N = (\Omega_2, P_2)$ là hai không gian xác suất, thì tích $\Omega_1 \times \Omega_2$, cũng

có một độ đo xác suất P, được xác định một cách tự nhiên bởi P_1 và P_2 bằng công thức sau: Nếu $A_1 \subset \Omega_1$ và $A_2 \subset \Omega_2$ nằm trong các sigma-đại số tương ứng của P_1 và P_2 thì:

$$P(A_1 \times A_2) = P_1(A_1) \times P_2(A_2). \tag{1.15}$$

Sigma-đại số của P chính là sigma đại số sinh bởi các tập con của $\Omega_1 \times \Omega_2$ có dạng $A_1 \times A_2$ như trên. Khi ta nói đến tích trực tiếp của hai không gian xác suất, ta sẽ hiểu là nó đi kèm độ đo xác suất được xác định như trên.

Tương tự như vậy, ta có thể định nghĩa tích trực tiếp của n không gian xác suất, hay thậm chí tích trực tiếp của một dãy vô hạn các không gian xác suất.

Định lý 1.2. *Hai phép chiếu tự nhiên từ tích* $(\Omega_1, P_1) \times (\Omega_2, P_2)$ *của hai không gian xác suất xuống* (Ω_1, P_1) *và* (Ω_2, P_2) *là hai ánh xạ bảo toàn xác suất.*

Ví dụ 1.13. Lấy 1 đồng xu tung 3 lần, mỗi lần hiện lên S (sấp) hoặc N (ngửa). Không gian xác suất các sự kiện ở đây là không gian các dãy 3 chữ cái mà mỗi chữ cái là S hay N:

$$\Omega = \{SSS, SSN, SNS, SNN, NSS, NSN, NNS, NNN\}.$$

Ký hiệu $(\Omega_k = \{S_k, N_k\}, P_k)$ là không gian xác suất của mặt hiện lên trong lần tung thứ k. Ta giả sử các kết quả của các lần tung là độc lập với nhau (tức là kết quả lần trước không ảnh hưởng đến kết quả của các lần sau), khi đó Ω có thể coi là tích trực tiếp của các không gian xác suất $(\Omega_k = \{S_k, N_k\}, P_k)$. Giả sử đồng xu là "cân bằng", hai mặt

sấp ngửa có xác suất hiện lên giống nhau trong mỗi lần tung. Khi đó các không gian $(\Omega_k = \{S_k, N_k\}, P_k)$ là đẳng cấu với nhau và với một không gian xác suất Bernoulli với tham số $p = 1/2$. Ta có thể viết: $\Omega = \{S, N\}^3$

Ví dụ 1.14. Trong ví dụ trên, nếu thay vì chỉ tung đồng xúc sắc có 3 lần, ta hình dùng la ta tung vô hạn lần (trong thực tế không làm được như vậy, nhưng cứ giả sử ta có vô hạn thời gian và làm được như vậy). Khi đó mỗi sự kiện được có thể được đánh dấu bằng một dãy vô hạn các chữ cái mà mỗi chữ là S hoặc N, và không gian xác suất là

$$\Omega = \{S, N\}^{\mathbb{N}}$$

Ta có thể xây dựng một ánh xạ bảo toàn xác suất sau từ $\{S, N\}^{\mathbb{N}}$ vào đoạn thẳng $[0, 1]$ với phân bố xác suất đều trên đó:

$$\phi((M_i)_{i\in\mathbb{N}}) := \sum_{i=1}^{\infty} \chi(M_i)/2^i$$

Ở đây mỗi M_i là S hoặc N, và $\chi(N) = 0, \chi(S) = 1$. Ánh xạ

$$\phi : \{S, N\}^{\mathbb{N}} \to [0, 1]$$

xây dựng như trên không phải là một song ánh, nhưng nó là một đẳng cấu xác suất!

Ví dụ 1.15. **Bài toán Méré**. Hiệp sĩ de Méré (tên khai sinh là Antoine Gombaud (1607-1684), là nhà văn và nhà triết học người Pháp) là một nhân vật lịch sử nghiện đánh bạc. Ông ta hay chơi xúc sắc, và nhận thấy rằng trong hai sự kiện sau:
A = “Tung một con xúc sắc 4 lần, có ít nhất 1 lần hiện lên 6”, và



Hình 1.5: Blaise Pascal (1623-1662)

$B =$ "Tung một đôi xúc sắc 24 lần, có ít nhất 1 lần hiện lên một đôi 6",

thì B ít xảy ra hơn A. Tuy nhiên ông ta không giải thích được tại sao. Theo ông ta thì đáng nhẽ hai sự kiện đó phải có khả năng xảy ra bằng nhau, vì $24 = 6 \times 4$. Ông ta bèn hỏi bạn mình là nhà toán học và triết học Blaise Pascal (1623-1662), vào năm 1654. Pascal lúc đó đã "từ bỏ toán", nhưng có nhận lời suy nghĩ về câu hỏi của de Méré. Sau đó Pascal viết thư trao đổi với Pierre de Fermat (159?-1665), một luật sư đồng thời là nhà toán học ở vùng Toulouse (Pháp). Hai người cùng nhau phát minh ra *lý thuyết xác suất cổ điển*, và giải được bài toán

của de Méré. Kết quả là: $P(A) = 1 - P(\overline{A}) = 1 - (1 - 1/6)^4 \approx 0,5177$, và $P(B) = 1 - P(\overline{B}) = 1 - (1 - (1/6)^2)^{24} \approx 0,4914$.



Hình 1.6: Fermat và "nàng toán". Tượng ở Toulouse.

Bài tập 1.7. Chứng minh định lý 1.2.

1.2.7 Phân bố nhị thức

Phân bố nhị thức là một trong những phân bố hay gặp nhất, và nó là một ví dụ về sự xuất hiện các phép toán tổ hợp trong xác suất thống kê.

Định nghĩa 1.4. Phân bố nhị thức *với các tham số* n, p *(*$n \in \mathbb{N}, 0 \leq p \leq 1$*) là phân bố xác suất*

$$P(k) = C_n^k p^k (1-p)^{n-k} \tag{1.16}$$

trên tập hợp $\Omega = \{0, 1, 2, \ldots, n\}$.

Ở đây, $C_n^k = \dfrac{n!}{k!(n-k)!}$ là nhị thức Newton. Ý nghĩa tổ hợp của C_n^k là: nó là số các tập con có đúng k phần tử trong một tập hợp có n phần tử, hay nói cách khác, nó là số cách chọn ra một nhóm con với k phần tử, từ một nhóm có n phần tử.

Nhắc lại rằng ta có công thức đại số quen thuộc sau:

$$(x+y)^n = \sum_{k=0}^{n} C_n^k x^k y^{n-k}. \tag{1.17}$$

Nếu thay x bằng p và y bằng $1-p$ trong công thức trên, thì ta có $\sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} = 1$, chứng tỏ định nghĩa phân bố xác suất nhị thức trên phù hợp với các tiên đề về xác suất.

Ý nghĩa của phân bố nhị thức như sau: Khi ta làm n lần một phép thử nào đó, và mỗi lần thì xác suất xảy ra kết quả A nào đó là p (ví dụ: một người bắn súng n lần, xác suất trúng đích mỗi lần là p), và giả sử là kết quả của các lần thử khác nhau độc lập với nhau (lần thử

này không ảnh hưởng đến lần thử khác), thì tổng số lần xảy ra kết quả A trong số n lần đó là một số nguyên nằm giữa 0 và n, và với mỗi $k = 0, 1, 2, \ldots, n$, xác sất của sự kiện "số lần ra kết quả A là k" bằng $C_n^k p^k (1-p)^{n-k}$.

Thật vậy, nếu ta lấy không gian xác suất cho mỗi phép thử là không gian $\{A, \overline{A}\}$, thì không gian xác suất các trường hợp của n lần thử là $\{A, \overline{A}\}^n$ (các phần tử của không gian này là các dãy n kết quả, mà mỗi kết quả là A hoặc $\overline{A}$. Có C_n^k phần tử của không gian $\{A, \overline{A}\}^n$ có chứa đúng k kết quả A và $(n-k)$ kết quả $\overline{A}$. Xác suất của mỗi phần tử đó là $p^k(1-p)^k$ theo công thức tích của xác suất. Bởi vậy xác suất của sự kiện "kết quả A xảy ra k lần" số phần tử của sự kiện này (hiểu như là một tập con của không gian xác suất) nhân với xác suất của một phần tử (vì các phần tử này có cùng xác suất), và bằng $C_n^k p^k (1-p)^{n-k}$.

Bài tập 1.8. Hai vận động viên Nam và Tiến chơi một trận tennis. Ai thắng được 3 set trước thì thắng cả trận. Giả sử xác suất để Nam thắng mỗi set là 40% (để Tiến thắng mỗi set là 60%, và kết quả của set này không ảnh hưởng đến set khác). Hỏi xác suất để Nam thắng trận tennis là bao nhiêu?

1.3 Xác suất có điều kiện

1.3.1 Định nghĩa xác suất có điều kiện

Như chúng ta đã biết, xác suất của một sự kiện có thể phụ thuộc vào nhiều yếu tố, điều kiện khác nhau. Để chỉ ra một cách cụ thể hơn

về việc xác suất của một sự kiện A nào đó phụ thuộc vào một điều kiện B nào đó ra sao, người ta đưa ra khái niệm xác suất có điều kiện. Điều kiện B cũng có thể hiểu là một sự kiện, tức là sự kiện "có B".

Định nghĩa 1.5. *Giả sử (trong một không gian xác suất nào đó) điều kiện B có xác suất khác không, $P(B) > 0$, thì* **xác suất của sự kiện** A **dưới điều kiện** B*, ký hiệu là $P(A|B)$, được định nghĩa như sau:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{1.18}$$

Một hệ quả trực tiếp của định nghĩa xác suất có điều kiện là công thức tích sau đây:

$$P(A \cap B) = P(A|B).P(B). \tag{1.19}$$

Tất nhiên, ta cũng có thể coi B là sự kiện, A là điều kiện, và khi đó ta có $P(A \cap B) = P(B|A).P(A)$

Ví dụ 1.16. Một lớp học có 30 bạn, trong đó có 17 bạn nữ và 13 bạn nam. Có 3 bạn tên là Thanh, trong đó có 1 bạn nữ và 2 bạn nam. Thầy giáo gọi ngẫu nhiên 1 bạn lên bảng. Xác suất để bạn đó có tên là Thanh sẽ là 1/10. Nhưng với điều kiện "đó là bạn nữ" thì xác suất để bạn đó tên là Thanh là 1/17. Sự kiện ở đây là $A =$ "tên là Thanh", và điều kiện là $B =$ "nữ". Không gian xác suất Ω có 30 phần tử, với phân bố xác suất đều. A có 3 phần tử, B có 17 phần tử, và $A \cap B$ có 1 phần tử. Bởi vậy: $P(A) = \frac{\#A}{\#\Omega} = 3/30 = 1/10$; $P(A|B) = P(A \cap B)/P(B) = (1/30)/(17/30) = 1/17$. Chú ý rằng, trong ví dụ này ta có $P(A|B) \neq P(A)$. Vẫn ví dụ này, nếu thầy giáo gọi 1 bạn có

tên là Thanh lên bảng, thì xác suất để bạn đó là bạn nữ là bao nhiêu? Lời giải: trong 3 bạn Thanh có 1 bạn là nữ, bởi vậy xác suất là 1/3. Sử dụng công thức $P(A \cap B) = P(B|A).P(A)$ với xác suất có điều kiện, ta cũng có $P(B|A) = P(A \cap B)/P(A) = (1/30)/(1/10) = 1/3$. (Câu hỏi: Vì sao hai cách giải khác nhau lại ra kết quả giống nhau?)

Ghi chú 1.3. Có thể giải thích ý nghĩa triết lý và toán học của định nghĩa xác suất có điều kiện như sau: Sự kiện A cùng với điều kiện B chính là sự kiện $A \cap B$, tức là "cả A và B cùng xảy ra". Ta có thể coi A và B là hai tập con của một không gian xác suất Ω ban đầu. Các tập con của B chính là các sự kiện với điều kiện B được thỏa mãn. Khi chúng ta đặt điều kiện B, thì tức là chúng ta đã hạn chế không gian xác suất từ Ω xuống còn B, và hạn chế các sự kiện A xuống còn $A \cap B$. Xác suất của A với điệu kiện B chính là xác suất của $A \cap B$ trong không gian xác suất mới B với một độ đo xác suất P_1: $P(A|B) = P_1(A \cap B)$. Độ đo xác suất P_1 không tùy ý, mà nó được sinh ra bởi độ đo xác suất P ban đầu, theo nguyên tắc "bình quân": nếu C và D là hai tập con của B (tức là 2 sự kiện thỏa mãn điều kiện B) với cùng xác suất, $P(C) = P(D)$, thì ta cũng phải coi rằng chúng có cùng xác suất có điều kiện: $P_1(C) = P_1(D)$. Một cách tổng quát hơn, ta có công thức tỷ lệ thuận: $P(C)/P(D) = P_1(C)/P_1(D)$ nếu C và D là hai tập con của B. Từ đó suy ra: $P(A \cap B)/P(B) = P_1(A \cap B)/P_1(B) = P_1(A \cap B) = P(A|B)$ (bởi vì $P_1(B) = 1$).

Ví dụ 1.17. Theo một con số thống kê ở Mỹ năm 2007, có khoảng 40% các vụ tai nạn xe cộ gây chết người là có người lái say rượu. Giá sử tỷ lệ số người say rượu khi lái xe là 4%. Hỏi việc xay rượu khi lái xe làm tăng khả năng gây tai nạn chết người lên bao nhiêu lần?

Nói cách khác, chúng ta muốn tính tỷ lệ $P(A|S)/P(A)$, ở đây A là sự kiện "lái xe xảy ra tai nạn chết người", S là điều kiện "người lái say rượu". Từ công thức $P(A \cap S) = P(A|S).P(S) = P(S|A).P(A)$ ta có $P(A|S)/P(A) = P(S|A)/P(S) = 40\%/4\% = 10$, tức là việc say rượu khi lái xe có thể làm tăng khả năng gây tai nạn xe cộ chết người lên 10 lần.

Bài tập 1.9. Có hai sự kiện A và B với xác suất lớn hơn 0. Khi nào thì ta có $P(A|B) = P(B|A)$?

Bài tập 1.10. Ta biết rằng một nhà nọ có 3 con mèo, trong đó có ít nhất 1 con là mèo cái. Hỏi rằng xác suất để cả 3 con mèo đều là mèo cái là bao nhiêu?

1.3.2 Sự độc lập và phụ thuộc của các sự kiện

Thế nào là hai sự kiện độc lập với nhau? Về mặt triết lý, hai sự kiện độc lập là hai sự kiện không liên quan gì đến nhau. Ví dụ, tôi không liên quan gì đến đội bóng đá Barcelona. Đội đó đá thắng hay thua tôi cũng không quan tâm, không ảnh hưởng gì đến việc tôi có phải đi chợ hay không. Hai sự kiện "tôi đi chợ" và "đội Barcelona thắng" có thể coi là độc lập với nhau. Nếu hai sự kiện A và B độc lập với nhau, thì việc có xảy ra hay không sự kiện B không ảnh hưởng gì đến việc có xảy ra hay không sự kiện A. Nói cách khác, xác suất của A với điều kiện B không khác gì xác suất của A khi không tính đến điều kiện B. Đấy chính là định nghĩa trong lý thuyết xác suất về sự độc lập của hai sự kiện:

Định nghĩa 1.6. *Sự kiện A được gọi là* **độc lập** *với sự kiện B nếu như*

$$P(A) = P(A|B) = P(A \cap B)/P(B), \tag{1.20}$$

hay viết cách khác:

$$P(A \cap B) = P(A).P(B) \tag{1.21}$$

Ghi chú 1.4. Công thức $P(A|B) = P(A)$ tương đương với công thức $P(A \cap B) = P(A).P(B)$ và tương đương với $P(B|A) = P(B)$. Điều đó có nghĩa là quan hệ *độc lập* là một quan hệ đối xứng: nếu A độc lập với B thì B độc lập với A, và chúng ta có thể nói là A và B độc lập với nhau. Trong công thức $P(A|B) = P(A)$ ta phải giả sử là $P(B) \neq 0$. Kể cả khi $P(B) = 0$ thì công thức $P(A \cap B) = P(A).P(B)$ vẫn có thể dùng làm định nghĩa được, và khi đó nó hiển nhiên đúng: một sự kiện có xác suất bằng 0 thì độc lập với mọi sự kiện khác.

Tổng quát hơn, giả sử ta có một họ $\mathcal{M}$ (hữu hạn hoặc vô hạn) các sự kiện.

Định nghĩa 1.7. *Họ $\mathcal{M}$ được gọi là một* **họ các sự kiện độc lập**, *nếu như với bất kỳ số tự nhiên k nào và bất kỳ k sự kiện $A_1, \ldots, A_k$ khác nhau nào trong họ $\mathcal{M}$ ta cũng có:*

$$P\left(\bigcap_{i=1}^{k} A_i\right) = \prod_{i=1}^{k} P(A_i). \tag{1.22}$$

Nếu như $P(A \cap B) = P(A).P(B)$ với bất kỳ hai sự kiện khác nhau nào trong họ $\mathcal{M}$ (tức là đẳng ta chỉ yêu cầu đẳng thức trên đúng trong trường hợp $k = 2$, thì họ $\mathcal{M}$ được gọi là họ các sự kiện độc lập từng đôi một.

Ghi chú 1.5. Tất nhiên nếu ta có một họ các sự kiện độc lập, thì các sự kiện trong họ độc lập từng đôi một với nhau. Nhưng điều ngược lại không đúng: Có những họ không độc lập, mà trong đó các sự kiện độc lập từng đôi một với nhau!

Ví dụ 1.18. Tung 1 xúc sắc 2 lần, được 2 số ký hiệu là a, b. Xét 3 sự kiện sau: X là sự kiện "$a + b$ là số chẵn", Y là sự kiện "$a = 1$" và Z là sự kiện "$b = 4$". Ở đây không gian xác suất là không gian có $6^2 = 36$ phần tử, mỗi phần tử là một cặp số (a, b), mỗi số có thể nhận 1 trong 6 giá trị 1,2,3,4,5,6. Ta có thể giả sử không gian xác suất này có phân bố xác suất đều (2 lần tung độc lập với nhau). Khi đó dễ dàng kiểm tra rằng các sự kiện X, Y, Z độc lập từng đôi một với nhau, thế nhưng họ 3 sự kiên $\{X, Y, Z\}$ không phải là một họ độc lập: $P(X \cap Y \cap Z) = 0$ trong khi $P(X).P(Y).P(Z) = (1/2).(1/6).(1/6) \neq 0$

Nếu như hai sự kiện không độc lập với nhau, thì người ta nói là chúng phụ thuộc vào nhau. Do tính chất đối xứng, nếu sự kiện A phụ thuộc vào sự kiện B thì B cũng phụ thuộc vào A. Nếu như $P(A|B) > P(B)$ thì ta có thể nói là điều kiện B *thuận lợi* cho sự kiện A, và ngược lại nếu $P(A) < P(A|B)$ thì điều kiện B *không thuận lợi* cho sự kiện A.

Công thức $P(A|B)P(B) = P(B|A)P(A)$ tương đương với công thức

$$P(A|B)/P(A) = P(B|A)/P(B), \tag{1.23}$$

có thể được suy diễn như sau: B thuận lợi cho A (tức là $P(A|B)/P(B) > 1$) thì A cũng thuận lợi cho B và ngược lại.

Ví dụ 1.19. Giả sử cứ 5 học sinh thì có 1 học sinh giỏi toán, cứ 3 học sinh thì có 1 học sinh giỏi ngoại ngữ, và trong số các học sinh giỏi

toán thì cứ 2 học sinh có 1 học sinh giỏi ngoại ngữ (lớn hơn tỷ lệ trung bình). Khi đó trong số các học sinh giỏi ngoại ngữ, tỷ lệ học sinh giỏi toán là 30% (cũng lớn hơn tỷ lệ trung bình): (1/2)/(1/3) = 30%/(1/5).

Bài tập 1.11. Chứng minh rằng nếu một sự kiện A độc lập với sự kiện B, thì nó cũng độc lập với sự kiện $\overline{B}$.

Bài tập 1.12. Tìm một ví dụ với 3 sự kiện A, B, C sao cho A độc lập với hai sự kiện B và C, nhưng không độc lập với $B \cap C$.

Bài tập 1.13. Lấy một bộ bài tú lơ khơ 52 quân, và rút ra từ đó 2 lần mỗi lần 1 quân, để được 2 quân. Gọi A là sự kiện "quân rút ra đầu tiên là quân nhép" và B là sự kiện "quân rút ra thứ hai là quân cơ". Hỏi hai sự kiện A và B có độc lập với nhau không?

1.3.3 Công thức xác suất toàn phần

Định nghĩa 1.8. *Một họ các tập con $B_1, \ldots, B_n$ của không gian xác suất Ω là một* **phân hoạch** *(partition) của Ω nếu như các tập B_i đôi một không giao nhau, và hợp của chúng bằng Ω:*

$$B_i \cap B_j = \emptyset \; \forall \; i \neq j \; , \; \cup_{i=1}^{n} B_i = \Omega. \qquad (1.24)$$

Nếu như ta chưa biết xác suất $P(A)$ của một sự kiện A nào đó, nhưng biết các xát suất $P(B_i)$ của một phân hoạch $(B_1, \ldots, B_n)$ của không gian xác suất, và biết các xác suất có điều kiện $P(A|B_i)$, thì ta có thể dùng công thức sau, gọi là **công thức xác suất toàn phần** (total probability formula), để tính xác suất của A:

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i).P(B_i). \qquad (1.25)$$

Trường hợp riêng của công thức trên là khi ta có hai sự kiện A, B, có thể sử dụng phân hoạch ($B, \overline{B} = \Omega \setminus B$) hai thành phần của Ω để tính xác suất của A:

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) = P(A|B).P(B) + P(A|\overline{B}).P(\overline{B}). \tag{1.26}$$

Bài tập 1.14. Theo một số liệu thống kê, năm 2004 ở Canada có 65,0% đàn ông là thừa cân[(1)], và 53,4% đàn bà thừa cân. Số đàn ông và đàn bà ở Canada coi như bằng nhau. Hỏi rằng, trong năm 2004, xác suất để một người Canada được chọn ngẫu nhiên là người thừa cân bằng bao nhiêu?

1.3.4 Công thức Bayes

Công thức Bayes, mang tên của linh mục và nhà toán học người Anh Thomas Bayes (1702-1761), là công thức ngược, cho phép tính xác suất có điều kiện $P(B|A)$ khi biết xác suất có điều kiện $P(A|B)$ và một số thông tin khác. Dạng đơn giản nhất của công thức này là: Nếu A, B là hai sự kiện bất kỳ với xác suất khác 0 thì ta có:

$$P(B|A) = \frac{P(A|B).P(B)}{P(A)}. \tag{1.27}$$

Công thức trên là hệ quả trực tiếp của công thức $P(B|A).P(A) = P(A|B).P(B) = P(A \cap B)$ đã được bàn đến ở những phần trước. Kết

[(1)] Theo định nghĩa của các tổ chức y tế, những người có chỉ số trọng lượng cơ thể (body mass index) ≥ 25 được gọi là thừa cân (overweight or obese), trên 30 được gọi là béo phì (obese), trên 40 là béo bệnh hoạn (morbidly obese). Chỉ số trọng lượng cơ thể được tính ra từ chiều cao và cân nặng theo công thức: BMI = trọng lượng (tính theo kg) chia cho chiều cao (tính theo mét) bình phương

hợp công thức trên với công thức xác suất toàn phần cho $P(A)$, ta được:

Định lý 1.3. *Giả sử* $(B_1, \ldots, B_n)$ *là một phân hoạch của không gian xác suất. Khi đó ta có* **công thức Bayes** *sau:*

$$P(B_k|A) = \frac{P(A|B_k).P(B_k)}{P(A)} = \frac{P(A|B_k).P(B_k)}{\sum_{i=1}^{n} P(A|B_i).P(B_i)}. \qquad (1.28)$$

với mọi $k = 1, 2, \ldots, n$.



Hình 1.7: Thomas Bayes (1702-1761)

Công thức Bayes rất đơn giản nhưng nó có ý nghĩa rất sâu xa. Một trong những lỗi mà rất nhiều người mắc phải, là lẫn lộn giữa $P(A|B)$ và $P(B|A)$, coi hai con số đó như là bằng nhau. Nhưng công thức

Bayes cho thấy hai con số đó có thể chênh lệch nhau rất nhiều, nếu như $P(A)$ và $P(B)$ chênh lệch nhau rất nhiều! Dưới đây là một ví dụ minh họa điều đó.

Ví dụ 1.20. Đây là một bài toán được 3 nhà toán học Cassels, Shoenberger và Grayboys đem đố 60 sinh viên và cán bộ y khoa tại Harvard Medical School năm 1978[(2)]. Giả sử có một loại bệnh mà tỷ lệ người mắc bệnh là 1/1000. Giả sử có một loại xét nghiệm, mà ai mắc bệnh khi xét cũng ra phản ứng dương tính, nhưng tỷ lệ phản ứng dương tính nhầm (false positive) là 5% (tức là trong số những người không bị bệnh có 5% số người thử ra phản ứng dương tính). Hỏi khi một người xét nghiệm bị phản ứng dương tính, thì khả năng mắc bệnh của người đó là bao nhiêu? Theo bạn là bao nhiêu? Hãy thử tự tìm câu trả lời trước khi đọc tiếp.

Nếu bạn trả lời 95% (= 100% - 5%), thì câu trả lời của bạn cũng giống câu trả lời của phần lớn những người khác được hỏi. Ta hãy thử phân tích kỹ thêm về câu hỏi này. Nếu ký hiệu K là sự kiên "không bị bệnh" và D là sự kiện phản ứng dương tính, thì con số 5% là con số $P(D|K)$ (xác suất có phản ứng dương tính khi mà không bị bệnh) chứ không phải $P(K|D)$ (xác suất không bị bệnh khi mà có phản ứng dương tính). Để tính $P(K|D)$, ta dùng công thức Bayes $P(K|D) = \dfrac{P(D|K).P(K)}{P(D|\bar{K}).P(\bar{K}) + P(D|K).P(K)}$. Ta có $P(D|K) = 5/100$, $P(K) = 1-1/1000 = 999/1000$, và $P(D|\bar{K}).P(\bar{K})+P(D|K).P(K) = (1).(1/1000) + (5/100).(999/1000) = 51/1000$ (tính xấp xỉ), và bởi vậy: $P(K|D) = (5/100).(999/1000)/(51/1000) \approx 98\%$. Như vậy trong

[(2)]Nguồn: Cassels, Schoenberger and Grayboys, Interpretation by physicians of clinical laboratory results. New England Journal of Medicine, 299 (1978), 999-1000.

số những người xét nghiệm ra dương tính, có khoảng 98% số người là không bị bệnh. Nói cách khác, khi xét nghiệm ra dương tính, xác suất để thực sự mắc bệnh chỉ có 2%!

Bài tập 1.15. Được biết có 5% đàn ông bị mù màu, và 0,25% đàn bà bị mù màu. Giả sử số đàn ông bằng số đàn bà. Chọn 1 người bị mù màu một cách ngẫu nhiên. Hỏi rằng xác suất để người đó là đàn ông là bao nhiêu?

1.4 Một số nghịch lý trong xác suất

Tính toán xác suất là một vấn đề nhiều khi hết sức tế nhị. Kể cả trong những bài toán tưởng chừng như rất đơn giản, cũng có thể tính ra kết quả sai mà khó phát hiện sai ở đâu. Phần này sẽ gồm một số "nghịch lý" trong xác suất để minh họa điều đó . Những nghịch lý này cho thấy chúng ta cần hết sức cẩn thận trong lúc lập mô hình tính toán xác suất, đặc biệt là xác suất có điều kiện, kiểm tra lại những điều tưởng chừng như hiển nhiên, để tránh sai lầm.

1.4.1 Nghịch lý 1 (Nghịch lý Simpson). Thuốc nào tốt hơn?

Một người nghiên cứu muốn xác định xem giữa 2 loại thuốc cùng để chữa 1 bệnh, loại nào tốt hơn. Kết quả thống kê về lượng người chữa được khỏi bệnh, phân biệt theo giới tính, được viết dưới đây

Giới tính: Nữ	Thuốc I	Thuốc II
Chữa được	150	15
Không chữa được	850	285

Giới tính: Nam	Thuốc I	Thuốc II
Chữa được	190	720
Không chữa được	10	180

Dựa vào bảng thống kê trên, có 2 câu trả lời trái ngược nhau như sau cho câu hỏi thuốc nào tốt hơn:
1) Thuốc I đem cho 1200 người dùng, chữ được bệnh cho 340 người. Thuốc II đem cho 1200 người dùng, chữa được 735 người, như vậy thuôc II tốt hơn.
2) Đối với nữ, tỷ lệ chữa được bệnh của Thuốc I là 15%, của Thuốc II là 5%. Đối với nam, tỷ lệ chữa được bệnh của thuốc I là 95%, của thuốc II là 80%. Trong cả hai trường hợp thì tỷ lệ chữa được bệnh của thuốc I cao hơn, vậy nên thuốc I tốt hơn.

Trong hai câu trả lời trên câu trả lời nào đáng tin? Vì sao? Nghịch lý nằm ở đâu?

1.4.2 Nghịch lý 2. Hoàng tử có chị em gái không?

Biết rằng cha mẹ của hoàng tử Romeo có 2 con (hoàng tử Romeo là một trong hai người con đó). Hỏi xác suất để hoàng tử Romeo có sister (chị gái hoặc em gái) là bao nhiêu? Có 2 đáp án sau:
1) Hoàng tử có 1 người anh chị em ruột. Có hai khả năng: hoặc người đó là con trai, hoặc là con gái. Như vậy xác suất để người đó là con gái (tức là hoàng tử có sister) là 1/2.
2) Có 4 khả năng cho 1 gia đình có 2 con: {B,B}, {B,G}, {G,B}, {G,G}. (B = boy = con trai, G = girl = con gái, xếp theo thứ tự con thứ nhất - con thứ hai). Vì ta biết hoàng tử là con trai (đây là điều kiện) nên

loại đi khả năng {G,G}, còn 3 khả năng {B,B}, {B,G}, {G,B}. Trong số 3 khả năng đó thì có 2 khả năng có con gái. Như vậy xác suất để hoàng tử có sister là 2/3.

Trong hai đáp án trên, ắt hẳn phải có (ít nhất) 1 đáp án sai. Thế nhưng cái nào sai, sai ở chỗ nào?

1.4.3 Nghịch lý 3. Văn Phạm có phải là thủ phạm?

Một người đàn ông tên là Văn Phạm bị tình nghi là thủ phạm trong một vụ án. Cảnh sát điều tra được những tin sau đây: 1) ngoài nạn nhân chỉ có 2 người có mặt lúc xảy ra vụ án, một trong hai người đó là Văn Phạm, người kia cảnh sát không hề biết là ai, và một trong hai người đó là thủ phạm; 2) thủ phạm phải là đàn ông. Hỏi xác suất để "Văn Phạm là thủ phạm" là bao nhiêu?

Gọi người thứ hai mà cảnh sát không biết là ai là "X". X có thể là đàn ông hoặc đàn bà. Ta gọi sự kiện "Văn Phạm là thủ phạm" là A, sự kiện "X là đàn ông" là B, "thủ phạm là đàn ông" là C. Có hai cách giải khác nhau như sau:

1) Theo công thức xác suất toàn phần ta có $P(A) = P(A|B).P(B) + P(A|\overline{B}).P(\overline{B})$ Nếu X là đàn bà thì X không thể là thủ phạm và Văn Phạm phải là thủ phạm, bởi vậy $P(A|\overline{B}) = 1$. Nếu X là đàn ông thì một trong hai người, X hoặc Văn Phạm, là thủ phạm, bởi vậy $P(A|B) = 1/2$. X có thể là đàn ông hoặc đàn bà, và ta coi số đàn ông bằng số đàn bà, bởi vậy $P(B) = P(\overline{B}) = 1/2$. Từ đó ta có $P(A) = (1/2).(1/2) + 1.(1/2) = 3/4$, có nghĩa là xác suất để "Văn Phạm là thủ phạm" bằng 3/4.

2) Ta coi C là điều kiện, và muốn tính xác suất có điều kiện $P(A|C)$ (xác suất để Văn Phạm là thủ phạm, khi biết rằng thủ phạm là đàn ông). Theo công thức Bayes ta có

$$P(A|C) = \frac{P(C|A).P(A)}{P(C|A).P(A) + P(C|\overline{A}).P(\overline{A})}.$$

Ở trong công thức trên, $P(A)$ là xác suất của sự kiện "Văn Phạm là thủ phạm" nếu như chưa có điều kiện "thủ phạm là đàn ông". Vì một trong hai người Văn Phạm và X là thủ phạm, nên xác suất $P(A)$ không có điều kiện ở đây là $P(A) = 1/2$. Ta có $P(C|A) = 1$ vì tất nhiên nếu Văn Phạm là thủ phạm thì thủ phạm là đàn ông. Ngược lại, $P(C|\overline{A}) = 1/2$ (nếu X là thủ phạm, thì thủ phạm có thể là đàn ông hoặc đàn bà, khi mà chưa đặt điều kiện "thủ phạm là đàn ông"). Bởi vậy ta có:

$$P(A|C) = \frac{1.(1/2)}{1.(1/2) + (1/2).(1/2)} = \frac{1/2}{3/4} = 2/3,$$

tức là xác suất để Văn Phạm là thủ phạm bằng 2/3.

Hai cách giải trên cho 2 đáp số khác nhau, như vậy (ít nhất) một trong hai cách giải trên là sai. Cách giải nào sai và sai ở chỗ nào?

1.4.4 Lời giải cho các nghịch lý

Nghịch lý 1. Vấn đề nằm ở chỗ Thuốc I được đem thử cho quá it nam, quá nhiều nữ so với thuốc II, nên khi lấy tổng số các kết quả của các phép thử thì nó thiên vị thuốc II và không phản ánh đúng tỷ lệ chữa được bệnh. Kết luận 1) là sai và kết luận 2) đáng tin hơn.

Nghịch lý 2. Nghịch lý này có trong 1 quyển giáo trình tiếng Anh về xác suất. Điều đáng ngạc nhiên là tác giả của giáo trình đó nói rằng đáp án thứ hai đúng (tức là xác suất = 2/3) và đáp án thứ nhất sai. Đọc kỹ đáp án thứ 2, ta thấy khả năng B,B thực ra không phải là một khả năng đơn, mà là một khả năng kép gồm có 2 khả năng trong đó: hoảng tử được nói đến hoặc là người con trai thứ nhất, hoặc là người con trai thứ hai. Như vậy phải tính B,B là 2 khả năng B=H,B và B, B=H (H là hoàng tử). Như thế tổng cộng vẫn có 4 khả năng, và xác suất vẫn là 2/4 = 1/2. Sai ở đây là sai trong cách đếm số khả năng. (Có câu hỏi khác: tại sao 4 khả năng này lại phải có xác suất bằng nhau? Tại sao lại phải có phân bố xác suất đều? Câu trả lời dành cho bạn đọc). Nếu ta đổi bài toán đi một chút thành: Một gia đình có 2 con, biết rằng ít nhất một trong hai con là con trai, thử hỏi xác suất để có con gái là bao nhiêu? Trong bài toán này thì xác suất là 2/3 thật. Bạn đọc thử nghĩ xem sự khác nhau giữa hai bài toán nằm ở chỗ nào?

Nghịch lý 3. Vấn đề ở đây nằm ở sự lẫn lỗn giữa các không gian xác suất trong lúc lập mô hình để tính xác suất. Trong cách giải thứ nhất, khi ta viết $P(A)$ để tính xác suất của sự kiện "Văn Phạm là thủ phạm", không gian xác suất của ta phải là không gian Ω_C tất cả các khả năng (với một trong 2 người Văn Phạm và X là thủ phạm) thỏa mãn điều kiện "thủ phạm là đàn ông", chứ không phải là không gian Ω của tất cả các khả năng có thể xảy ra (với một trong 2 người Văn Phạm và X là thủ phạm), bất kể thủ phạm là đàn ông hay đàn bà. Để cho khỏi lẫn lộn, thì trong cách giải thứ nhất ta phải viết $P_C(A) = P_C(A|B).P_C(B) + P_C(A|\overline{B}).P_C(\overline{B})$ Trong không gian Ω thì

ta có $P(B) = 1/2$, tức là xác suất để X là đàn ông là 1/2. Nhưng trong không gian Ω_C dùng trong cách giải thứ nhất, thì ta phải dùng xác suất P_C của không gian đó, và $P_C(B)$ không phải là 1/2, mà thực ra là 2/3, và $P_C(\overline{B}) = 1/3$. Nói cách khác, khi biết rằng một trong hai người X và Văn Phạm là thủ phạm, và biết rằng thủ phạm là đàn ông, thì xác suất để X là đàn ông là 2/3 chứ không còn là 1/2 nữa! (Vì sao vậy?). Nếu ta sử dụng các con số xác suất này trong công thức tính xác suất toàn phần của A trong không gian Ω_C thì ta được: $p_C(A) = (1/2).(2/3) + 1.(1/3) = 2/3$ Tức là nếu ta sửa lỗi về xác suất của B đi, thì cách giải thứ nhất sẽ cho cùng đáp số 2/3 như cách giải thứ hai.

1.5 Luật số lớn

Luật số lớn là một trong những định luật cơ bản nhất của lý thuyết xác suất và thống kê. Ở dạng đơn giản nhất, nó có thể được phát biểu một cách nôm na như sau: khi một phép thử được lặp đi lặp lại rất nhiều lần, thì số lần cho ra một kết quả nào đó trong tổng số các lần thử sẽ phản ánh khá chính xác xác suất để xảy ra kết quả đó trong 1 lần thử. Ví dụ, giả sử ta có một đồng tiền với hai mặt sấp (S) và ngửa (N) với xác suất hiện lên bằng nhau và bằng 1/2 khi tung đồng tiền. Giả sử ta tung đi tung lại đồng tiền nhiều lần, và được một dãy các kết quả sấp ngửa, ví dụ như: S N S S N S N S N N S S ... Ta gọi $S(n)$ là tần số xuất hiện lên mặt sấp sau khi tung đồng tiền n lần, tức là số lần hiện lên mặt sấp sau khi tung đồng tiền n lần chia cho n, ví dụ như theo dãy trên:

$S(1) = 1, S(2) = 1/2, S(3) = 2/3, S(4) = 3/4, S(5) = 3/5, S(6) = 2/3, S(7) = 4/7, S(8) = 5/8, S(9) = 5/9, S(10) = 1/2, S(11) = 6/11, S(12) = 7/12, \ldots$. Các con số $S(n)$ mà chúng ta thu được nói chung khác 1/2, nhưng luật số lớn nói rằng chúng ta có thể yên tâm rằng khi n tiến tới vô cùng thì $S(n)$ sẽ tiến tới 1/2: $\lim_{n\to\infty} S(n) = 1/2$.

Dưới đây chúng ta sẽ phát biểu luật số lớn một cách chặt chẽ thành định lý toán học và chứng minh nó, cho phân bố Bernoulli.

Giả sử có một phép thử nào đó có thể thực hiện được nhiều lần, và xác suất để xảy ra kết quả X trong một lần thử là một hằng số p, $0 < p < 1$. (Ví dụ: phép thử là "tung xúc sắc", kết quả là "hiện lên 1 chấm", xác suất là p=1/6). Ta gọi $X_{k,n}$ là sự kiện sau: khi thực hiện n lần phép thử thì X xuất hiện k lần trong số n lần thử. Chúng ta biết rằng xác suất của $X_{k,n}$ tuân theo phân bố nhị thức:

$$P(X_{k,n}) = C_n^k p^k (1-p)^{n-k}. \quad (1.29)$$

Lấy một số dương $\epsilon > 0$ tùy ý sao cho $0 < p - \epsilon < p + \epsilon < 1$. Gọi X_n^ϵ là sự kiện sau: khi làm phép thử n lần thì tần suất xuất hiện kết quả X chênh lệch so với xác suất p không quá ϵ, tức là $p - \epsilon \leq k/n \leq p + \epsilon$, trong đó k là số lần hiện lên kết quả X. Sự kiện X_n^ϵ là hợp của các sự kiện $X_{k,n}$ thỏa mãn bất đẳng thức $p - \epsilon \leq k/n \leq p + \epsilon$, do vậy:

$$P(X_n^\epsilon) = \sum_{n(p-\epsilon)\leq k\leq n(p+\epsilon)} C_n^k p^k (1-p)^{n-k}. \quad (1.30)$$

Định lý 1.4. *Với hai số dương p, ϵ bất kỳ thỏa mãn $0 < p - \epsilon < p + \epsilon < 1$, ta có*

$$\lim_{n\to\infty} \sum_{n(p-\epsilon)\leq k\leq n(p+\epsilon)} C_n^k p^k (1-p)^{n-k} = 1. \quad (1.31)$$

Có nghĩa là, xác suất $P(X_n^\epsilon)$ của sự kiện "sau n phép thử thì tần suất hiện kết quả X sai lệch so với xác suất p của X không quá ϵ" tiến tới 1 khi số phép thử n tiến tới vô cùng.

Định lý trên gọi là **dạng yếu của luật số lớn** cho phân bố Bernoulli. Dạng mạnh của luật số lớn, sẽ được xét tới trong chương sau, phát biểu là tần suất k/n tiến tới p khi n tiến tới vô cùng *hầu như chắc chắn* (tức là với xác suất bằng 1: tập những dãy vô hạn lần thử mà điều đó sai có xác suất bằng 0 trong không gian tất cả các dãy vô hạn lần thử).

Chứng minh. Chúng ta muốn chứng minh rằng hiệu

$$1 - P(X_n^\epsilon) = U_n + V_n, \tag{1.32}$$

trong đó

$$U_n = \sum_{0 \le k < n(p-\epsilon)} C_n^k p^k (1-p)^{n-k} \text{ và } V_n = \sum_{n(p+\epsilon) < k \le n} C_n^k p^k (1-p)^{n-k}, \tag{1.33}$$

tiến tới 0 khi n tiến tới vô cùng. Để đánh giá V_n, chúng ta có thể dùng thủ thuật sau đây: Gọi λ là một số dương bất kỳ, khi đó ta có

$$\begin{aligned} V_n &\le \sum_{n(p+\epsilon) < k \le n} e^{\lambda(k-n(p+\epsilon))} C_n^k p^k (1-p)^{n-k} \\ &\le \textstyle\sum_{0 \le k \le n} e^{\lambda(k-n(p+\epsilon))} C_n^k p^k (1-p)^{n-k} \\ &= e^{-\lambda n\epsilon} \sum_{0 \le k \le n} C_n^k (e^{\lambda(1-p)} p)^k (e^{-\lambda p}(1-p))^{n-k} \\ &= e^{-\lambda n\epsilon} (e^{\lambda(1-p)} p + e^{-\lambda p}(1-p))^n \\ &= [e^{-\lambda\epsilon} (e^{\lambda(1-p)} p + e^{-\lambda p}(1-p))]^n = f(\lambda)^n, \end{aligned} \tag{1.34}$$

với $f(\lambda) = e^{-\lambda\epsilon}(e^{\lambda(1-p)}p + e^{-\lambda p}(1-p))$. Chú ý rằng hàm số $f(\lambda)$ có $f(0) = 1$ và đạo hàm $f'(0) = -\epsilon < 0$. Như vậy nếu ta chọn $\lambda > 0$ đủ

nhỏ thì ta có $0 < f(\lambda) < 1$, dẫn tới $\lim_{n\to\infty} f(\lambda)^n = 0$. Vì $V_n \leq f(\lambda)^n$ nên ta cũng có $\lim_{n\to\infty} V_n = 0$. Một cách hoàn toàn tương tự, ta có thể chứng minh $\lim_{n\to\infty} U_n = 0$, suy ra $\lim_{n\to\infty} U_n + V_n = 0$. Phần chứng minh này là bài tập dành cho bạn đọc. □

Chúng ta có thể mở rộng luật số lớn cho phân bố Bernoulli thành luật số lớn cho một không gian xác suất bất kỳ vời hữu hạn các phần tử như sau. Giả sử có một phép thử, mà cứ một lần thử thì hiện lên một trong các kết quả $A_1, \ldots, A_s$, với các xác suất $P(A_i) = p_i$ tương ứng. ($\sum_{i=1}^s p_i = 1$, và $\Omega = \{A_1, \ldots, A_s\}$ lập thành một không gian xác suất hữu hạn với các xác suất $P(A_i) = p_i$). Làm phép thử đó n lần (các lần thử độc lập với nhau), và gọi k_i là số lần hiện lên kết quả A_i trong số n lần thử đó. Gọi $B^\epsilon_{n,i}$ là sự kiện $|\frac{k_i}{n} - p_i| < \epsilon$.

Định lý 1.5. *Với mọi $\epsilon > 0$ ta có*

$$\lim_{n\to\infty} P(B^\epsilon_{n,1} \cap B^\epsilon_{n,2} \cap \ldots \cap B^\epsilon_{n,s}) = 1. \tag{1.35}$$

Ghi chú 1.6. (Một chút lịch sử[3]). Luật số lớn được biết đến ở dạng trực giác, "càng thí nghiệm nhiều lần thì kết quả thống kê càng chính xác", từ hàng nghìn năm trước đây. Nhà toán học và thiên văn học người Ấn Độ Brahmagupta (598-668), và sau đó nhà toán học người Italia Gerolamo Cardano (1501-1576), có phát biểu nó mà không chứng minh. Người đầu tiên đưa ra chứng minh toán học cho luật số lớn có lẽ là Jacob Bernoulli năm 1713, và luật số lớn còn được gọi là Định lý Bernoulli. Cái tên *luật số lớn* (la loi des grands nombres) được Siméon Denis Poisson viết ra năm 1835, và ngày nay người ta hay gọi theo tên đó.

[3] Xem: http://en.wikipedia.org/wiki/Law_of_large_numbers.

Bài tập 1.16. Suy ra định lý 1.5 từ định lý 1.4.

1.6 Bài tập bổ sung cho Chương 1

Bài tập 1.17. Tung một đồng tiền cân bằng cho đến khi mặt ngửa hiện lên 3 lần. Gọi A là sự kiện "cần tung 6 lần". Hãy lập một không gian xác suất cho vấn đề xác suất này, và tính xác suất của sự kiện A.

Bài tập 1.18. (Bài tập của ngành bảo hiểm). Một công ty bảo hiểm ô tô có 20000 người đăng ký bảo hiểm. Những người đăng ký bảo hiểm được công ty phân loại theo 3 tiêu chuẩn:
i) Trẻ hay già,
ii) Đàn ông hay đàn bà.
iii) Có vợ/chồng hay độc thân.
Được biết, trong số những người đăng ký bảo hiểm, có 6300 người trẻ, 9600 người là đàn ông, 13800 người có vợ/chồng, 2700 đàn ông trẻ, 6400 đàn ông có vợ, 2900 người trẻ có vợ/chồng, 1100 người là đàn ông trẻ có vợ. Hỏi xác suất để một người đăng ký bảo hiểm ô tô của hãng được chọn một cách ngẫu nhiên là một phụ nữ trẻ độc thân bằng bao nhiêu?

Bài tập 1.19. Một anh chàng có 2 cô bạn gái A và B, và không biết là thích cô nào hơn. Anh ta hay đi thăm các cô bạn một cách ngẫu nhiên: ra bến xe buýt, nếu gặp xe buýt đi tuyến đường đến nhà cô A trước thì đi lên xe đó thăm cô A, còn nếu gặp xe đi tuyến đường đến nhà cô B trước thì đi thăm cô B. Cả hai tuyến đường đều có xe đều đặn 10 phút một xe. Sau một thời gian dài, anh ta nhận ra rằng mình đi thăm cô bạn A nhiều gấp 3 lần cô bạn B. Có thể giải thích bằng

xác suất tại sao?

Bài tập 1.20. (Số may rủi). Giả sử có một loại xổ số chỉ có 100 số, từ 00 đến 99, mỗi lần quay có 1 số trúng giải.
i) Tính xác suất sao cho trong 100 lần quay, không có lần nào số 68 trúng giải.
ii) Tính xác suất để sao cho trong 100 lần quay, số 99 trúng giải đúng 2 lần.

Bài tập 1.21. Một lớp học có 36 học sinh. Hỏi rằng xác suất để có hai học sinh của lớp có cùng ngày sinh nhật là bao nhiêu? (Viết công thức để tính số đó, và thử ước lượng xem số đó gần số nào hơn trong 3 số này: 0, 50%, 1?)

Ví dụ 1.21. Có n người chơi trò tung mũng trong một dạ hội: mỗi người cầm 1 cái mũ của mình, tung vào giữa phòng. Sau đó mỗi người nhặt lấy một cái mũ trong số các mũ được tung một cách ngẫu nhiên. Chứng minh rằng xác suất để không có người nào nhặt được đúng mũ của chính mình là

$$\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \ldots + \frac{(-1)^n}{n!}.$$

Khi n tiến tới vô cùng thì số này tiến tới e^{-1}.

Bài tập 1.22. (Bổ đề Borel–Cantelli). Giả sử $(A_n)_{n\in\mathbb{N}}$ là một dãy các tập con đo được trong một không gian xác suất (Ω, P). Gọi B_∞ là tập hợp các phần tử của Ω mà nằm trong một số vô hạn các tập con A_n của dãy. Chứng minh rằng:
i) Nếu $\sum_{n=1}^{\infty} P(A_n) < \infty$ thì $P(B_\infty) = 0$.
ii) Nếu tồn tại một số ϵ và vô hạn các tập con A_n của dãy thỏa mãn điều kiện $P(A_n) \geq \epsilon$, thì $P(B_\infty) \geq \epsilon$.

(Gợi ý: Đặt B_k = tập các phần tử của Ω nằm trong ít nhất k tập con A_n của dãy. Khi đó $P(B_\infty) = \lim_{k\to\infty} B_k$. Trong trường hợp thứ nhất, chứng minh rằng $k.B_k \leq \sum_{n=1}^{\infty} P(A_n)$ với mọi k. Trong trường hợp thứ hai, chứng minh rằng $P(B_k) \geq \epsilon$ với mọi k).

Bài tập 1.23. (Tủ của Bertrand). Có 3 ngăn kéo, 1 ngăn có 2 đồng tiền vàng, 1 ngăn có 2 đồng tiền bạc, và 1 ngăn có 1 đồng tiền vàng và 1 đồng tiền bạc. Rút ra một ngăn kéo một cách ngẫu nhiên, và lôi ra từ ngăn kéo đó một đồng tiền một cách ngẫu nhiên. Giả sử được 1 đồng tiền vàng. Hỏi xác suất để ngăn kéo được rút ra là ngăn kéo chứa hai đồng tiền vàng bằng bao nhiêu?

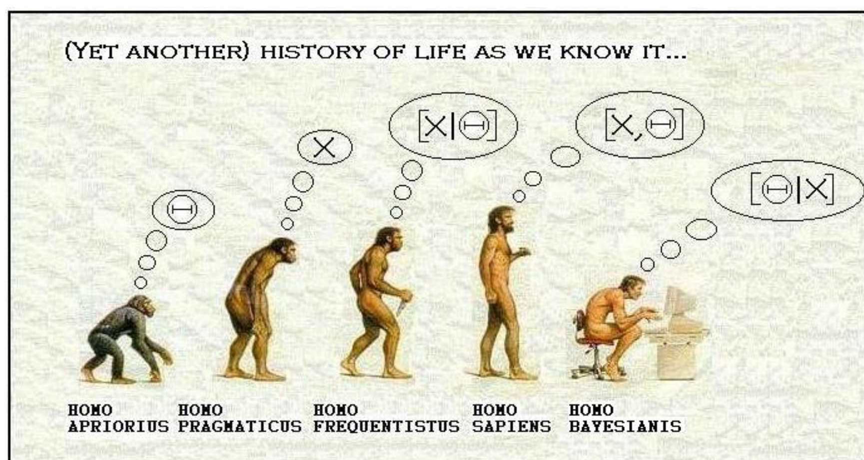
Bài tập 1.24. Có ba người A, B, C bị bắt vào tù. Có lệnh thả hai trong số ba người này ra. Cai tù nhận được lệnh, nhưng đến hôm sau mới được công bố và thi hành lệnh. Người tù A bảo cai tù: hãy nói cho tôi biết tên 1 người được thả trong hai người B và C đi. Cai ngục trả lời: anh đang có xác suất được thả là 2/3. Nếu tôi nói tên một người được thả trong số hai người B và C, thì giữa anh và người còn lại chỉ còn một người được thả nữa thôi, bởi vậy xác suất để anh được thả sẽ giảm xuống còn 1/2. Tôi không muốn xác suất để anh được thả bị giảm đi, bởi vậy tôi sẽ không nói tên. Hỏi rằng người cai ngục lý luận như vậy có đúng không?

Bài tập 1.25. Hai kẻ trộm đeo mặt nạ, bị cảnh sát đuổi bắt, bèn vứt mặt nạ đi và trà trộn vào một đám đông. Cảnh sát bắt giữ toàn bộ đám đông, tổng cộng 60 người, và dùng máy phát hiện nói dối (lie detector) để điều tra xem ai trong đám đông là kẻ trộm. Biết rằng đối với kẻ trộm, xác suất bị máy nghi là có tội là 85%, nhưng đối với người vô tội, thì xác suất để bị máy nghi nhầm thành có tội là 7%.

Giả sử X là một nhân vật trong đám đông bị máy nghi là có tội. Tính xác suất để X là kẻ trộm.

Bài tập 1.26. (Bò điên). Năm 2001 Cộng Đồng Châu Âu có làm một đợt kiểm tra rất rộng rãi các con bò để phát hiện những con bị *bệnh bò điên* (bovine spongiform encephalopathy). Không có xét nghiệm nào cho kết quả chính xác 100%. Một loại xét nghiệm, mà ở đây ta gọi là xét nghiệm A, cho kết quả như sau: khi con bò bị bệnh bò điên, thì xác suất để ra phản ứng dương tính trong xét nghiệm A là 70%, còn khi con bò không bị bệnh, thì xác suất để xảy ra phản ứng dương tính trong xét nghiệm A là 10%. Biết rằng tỷ lệ bò bị mắc bệnh bò điên ở Hà Lan là 1,3 con trên 100000 con. Hỏi rằng khi một con bò ở Hà Lan phản ứng dương tính với xét nghiệm A, thì xác suất để nó bị mắc bệnh bò điên là bao nhiêu?

Bài tập 1.27. (Giá dầu hỏa). Giá dầu hỏa có những lúc dao động rất mạnh, có khi đi lên hơn 100% trong vòng 1 năm. Giả sử rằng, nếu tính giá theo USD của năm 2009 (sau khi đã điểu chỉnh theo tỷ lệ lạm phát), thì giá dầu hỏa không bao giờ xuống dưới 10 USD một thùng (dưới mức đó người ta ngừng sản xuất dầu hỏa vì không còn lãi gì nữa) và không bao giờ lên quá 300 USD một thùng (trên mức đó người ta dùng các loại năng lượng khác rẻ hơn). Hỏi họ các sự kiện Gx sau đây (x=0,1,...9) có thể là một họ độc lập các sự kiện được không : Gx = “năm 201x giá dầu hỏa tăng lên ít nhất 50% tính từ đầu năm đến cuối năm, tính theo USD của năm 2009”. Giải thích tại sao?



Hình 1.8: Tranh vui về sự tiến hóa của loài người

Chương 2

Biến Ngẫu Nhiên

2.1 Biến ngẫu nhiên và phân bố xác suất của nó

2.1.1 Biến ngẫu nhiên là gì?

"Biến" là cái có thể thay đổi. "Ngẫu nhiên" là khi người ta chưa xác định được cái gì đó, thì người ta gọi nó là ngẫu nhiên. Cái gì khi đã xác định được, thì thành "định tính", hết ngẫu nhiên. Một biến có thể là ngẫu nhiên với người này, nhưng không ngẫu nhiên với người khác, tùy theo lượng thông tin nhận được. Ví dụ, số thứ tiếng ngoại ngữ mà ông A nói được là một số xác định, không ngẫu nhiên đối với ông A, nhưng nó là một số không xác định, ngẫu nhiên với một ông B nào đó.

Biến ngẫu nhiên có thể nhận giá trị trong mọi phạm trù (hiểu từ phạm trù ở đây theo nghĩa thông thường chứ không phải theo nghĩa phạm trù toán học), ví dụ như màu sắc, hình dạng, phương hướng,

v.v. Tuy nhiên, bằng các ánh xạ (không ngẫu nhiên), chúng ta có thể chuyển việc nghiên cứu mọi biến ngẫu nhiên về việc nghiên cứu các biến ngẫu nhiên nhận giá trị là các số. Bởi vậy ở đây, khi nói đến một biến ngẫu nhiên mà không nói cụ thể nó nhận giá trị ở đâu, chúng ta sẽ hiểu là các giá trị của nó là các con số.

Ví dụ 2.1. Tại thời điểm đóng cửa thị trường chứng khoán Mỹ hôm 04/09/2009, giá cổ phiếu của hãng phần mềm máy tính Oracle (mã chứng khoán: ORCL) là 21,97 USD. Nó đã được xác định và không còn ngẫu nhiên. Thế nhưng tại thời điểm đó, thì giá cố phiếu của Oracle cho lúc cuối ngày 18/09/2009 chưa được biết, và nó là một biết ngẫu nhiên đối với thị trường chứng khoán. Người ta cho rằng giá của nó vào ngày 18/09/2009 có thể lên trên 23 USD, mà cũng có thể xuống dưới 21 USD. Điều này thể hiện qua việc, tại thời điểm cuối ngày 04/09/2009 , quyền mua ORCL trước ngày 19/09/2009 với giá 23 USD (September 2009 call option at strike price 23) có giá 0,25 USD (nếu như ai cũng biết chắc rằng giá của ORCL vào thời điểm 18/09/2009 sẽ không vượt quá 23 thì cái quyền mua đó sẽ phải có giá bằng 0 vì không có giá trị gì), đồng thời quyền bán (put option) ORCL với giá 21 có giá là 0,30 USD. (Các thông tin về giá cả cổ phiếu và option có thể xem trên rất nhiều các trang web về chứng khoán).

Tương tự như với các số và các hàm số, ta có thể làm nhiều phép toán khác nhau với các biến ngẫu nhiên: cộng, trừ, nhân, chia, lấy giới hạn, tích phân, hàm hợp, v.v. Qua các phép toán như vậy, chúng ta có thể sinh ra các biến ngẫu nhiên mới từ các biến ngẫu nhiên cho trước.

Ví dụ 2.2. Một học sinh thi vào đại học phải thi 3 môn. Điểm của mỗi

môn có thể coi là 1 biến ngẫu nhiên. Tổng số điểm cũng là một biến ngẫu nhiên, và nó là tổng của 3 biến ngẫu nhiên phía trước.

Ví dụ 2.3. Tốc độ V của một xe ô tô đang chạy trên đường có thể coi là một biến ngẫu nhiên. Nếu xe đang chạy mà phải phanh gấp lại vì phía trước có nguy hiểm, thì từ thời điểm người lái xe bóp phanh cho đến thời điểm xe dừng lại, xe phải chạy thêm mất một quãng đường có độ dài D nữa. D cũng có thể coi là một biến ngẫu nhiên. Nó không phải là tỷ lệ thuận với V, mà là tỷ lệ thuận với bình phương của V. Tức là biến ngẫu nhiên D có thể được sinh ra từ biến ngẫu nhiên V theo công thức: $D = k.V^2$. Hệ số k ở đây phụ thuộc vào điều kiện của đường và điều kiện của xe; nó có thể coi là xác định nếu ta biết các điều kiện này, còn nếu không thì có thể coi là một biến ngẫu nhiên khác. Ví dụ, trong điều kiện bình thường, thì $k = 0,08m^{-1}.s^2$: một xe đang chạy với tốc độ 36km/h = 10m/s thì từ lúc bóp phanh đến lúc dừng lại chạy thêm mất $0,08 \times 10^2 = 8$ mét, nhưng nếu xe đang chạy với tốc độ $108km/h = 3 \times 36km/h$, thì từ lúc bóp phanh đến lúc dừng lại sẽ chạy thêm mất những $8 \times 3^2 = 72$ mét.

2.1.2 Mô hình toán học của biến ngẫu nhiên

Giả sử có một biến ngẫu nhiên X. Chúng ta giả sử là có nhiều tình huống khác nhau có thể xảy ra, và trong mỗi tình huống thì X sẽ nhận được một giá trị nào đó. Như vậy một biến ngẫu nhiên có thể được mô hình hóa bằng một hàm số $X : \Omega \to \mathbb{R}$. Ở đây Ω là không gian đại diện cho các tình huống có thể xảy ra. Các tình huống, hay các nhóm các tình huống (các tập hợp con của Ω) là các sự kiện, và chúng ta có thể gắn cho mỗi sự kiện một xác suất về khả năng xẩy

ra. Điều đó có nghĩa là Ω có thể coi là một không gian xác suất, ký hiệu là (Ω, P) với một độ đo xác suất P. Chúng ta luôn giả sử rằng, với mọi cặp số $a, b \in \mathbb{R}, a < b$, tồn tại xác suất $P(a < X \leq b)$ của sự kiện $(a < X \leq b)$, hay nói cách khác, tập hợp $\{\omega \in \Omega | a < X(\omega) \leq b\}$ là tập đo được. Các hàm $X : \Omega \to \mathbb{R}$ thỏa mãn điều kiện này được gọi là **hàm đo được** trên (Ω, P). Từ đó chúng ta có định nghĩa toán học sau:

Định nghĩa 2.1. *Một* **biến ngẫu nhiên** *(random variable) với giá trị thực là một hàm số đo được trên một không gian xác suất:*

$$X : (\Omega, P) \to \mathbb{R}. \tag{2.1}$$

Định nghĩa 2.2. *Nếu ta có hai biến ngẫu nhiên X, Y (với cùng một mô hình không gian xác suất), thì ta sẽ nói rằng $X = Y$ theo nghĩa xác suất, hay $X = Y$* **hầu khắp mọi nơi** *, nếu như sự kiện "$X = Y$" có xác suất bằng 1 (tức là tập hợp các trường hợp mà ở đó $X \neq Y$ có xác suất bằng 0, có thể bỏ qua).*

Ví dụ 2.4. Một thí sinh đi kiểm tra trắc nghiệm, được giao 5 câu hỏi một cách ngẫu nhiên. Được biết 3 câu đầu thuộc loại vừa, và xác suất để thí sinh làm đúng cho mỗi câu là 80%, 2 câu sau thuộc loại khó, và xác suất làm đúng mỗi câu là 50%. Mỗi câu làm đúng thì được tính 1 điểm. Không gian Ω các tình huống ở đây gồm $2^5 = 32$ phần tử, mỗi phần tử có thể được ký hiệu bằng 1 dãy 5 chữ cái mà mỗi chứ cái là D (đúng) hoặc S (sai). Từ thông tin phía trên có thể suy ra xác suất của mỗi phần tử của Ω, ví dụ như $P(DDSSD) = 80\%.80\%.20\%.50\%.50\% = 4/125 = 3,2\%$. Biến ngẫu nhiên là tổng số

điểm, tức là hàm $X : \Omega \to \{0, 1, 2, 3, 4, 5\}$, X của một dãy chữ cái bằng số lần chữ cái D xuất hiện trong dãy.

Ví dụ 2.5. Nếu A là một sự kiện, thì ta có thể định nghĩa **hàm chỉ báo** χ_A của A như sau: $\chi_A = 1$ khi A xảy ra và $\chi_A = 0$ khi A không xảy ra. Nếu ta có một sự kiện, thì hàm chỉ báo của nó là một biến ngẫu nhiên chỉ nhận hai giá trị 0 và 1, và ngược lại, nếu ta có một biến ngẫu nhiên F chỉ nhận 2 giá trị 0 và 1, thì nó là hàm chỉ báo của sự kiện $\{F = 1\}$. Nếu ta biểu diễn A như là một tập con của một không gian xác suất Ω, thì hàm chỉ báo của A được biểu diễn như là hàm chỉ báo của tập A trong Ω:

$$\chi_A(\omega) = \begin{cases} 1 & \text{khi } \omega \in A \\ 0 & \text{khi } \omega \in \overline{A} = \Omega \setminus A \end{cases} . \tag{2.2}$$

2.1.3 Phân bố xác suất của biến ngẫu nhiên

Nhắc lại rằng, nếu ta có một không gian xác suất (Ω, P) và một ánh xạ $X : (\Omega, P) \to \Lambda$ từ Ω lên một không gian Λ nào đó, thì phép push-forward theo X sẽ biến Λ thành một không gian xác suất, với độ đo xác suất cảm sinh $P_X = X^*P$: theo định nghĩa, nếu B là một tập con của Λ sao cho tồn tại $P(X^{-1}(B))$ thì

$$P_X(B) = P(X^{-1}(B))$$

Trong trường hợp $X : (\Omega, P) \to \mathbb{R}$ là một biến ngẫu nhiên, tính chất đo được của X (trong định nghĩa của biến ngẫu nhiên) nói rằng tồn tại $P(X^{-1}(]a, b])) = P(a < F \leq b)$ với mọi đoạn thẳng nửa mở $]a, b]$ trên $\mathbb{R}$. Sigma-đại số $\mathcal{B}$ sinh bởi các đoạn thẳng nửa mở trên $\mathbb{R}$ được

gọi là **signma-đại số Borel** của $\mathbb{R}$. Khi nói đến một phân bố xác suất trên $\mathbb{R}$, chúng ta sẽ coi rằng sigma-đại số tương ứng chính là sigma-đại số Borel, bởi vì nói chung chúng ta sẽ chỉ quan tâm đến xác suất của các đoạn thẳng, và các tập con của $\mathbb{R}$ xây dựng được từ các đoạn thẳng bằng các phép giao, hợp, lấy phần bù. Do đó ta có định nghĩa sau:

Định nghĩa 2.3. Phân bố xác suất *(hay còn gọi là* **phân phối xác suất***) của một biến ngẫu nhiên X (trên $\mathbb{R}$) là phân bố xác suất P_X trên $\mathbb{R}$, với sigma-đại số là sigma-đại số Borel $\mathcal{B}$ của $\mathbb{R}$, cho bởi công thức sau:*

$$P_F(B) = P(X^{-1}(B)) \tag{2.3}$$

với mọi tập con B của $\mathbb{R}$ nằm trong sigma-đại số $\mathcal{B}$.

Định lý sau cho phép hiểu rõ hơn về sigma-đại số Borel:

Định lý 2.1. *i) Mọi đoạn thẳng mở (bị chặn hay không bị chặn) đều là phần tử của sigma-đại số Borel. Ngược lại, sigma-đại số sinh bởi các đoạn thẳng mở cũng chính bằng sigma-đại số Borel.*
ii) Mọi đoạn thẳng đóng đều là phần tử của sigma-đại số Borel. Ngược lại, sigma-đại số sinh bởi các đoạn thẳng đóng cũng chính bằng sigma-đại số Borel.

Chứng minh. Giả sử $]a, b[$ là một đoạn thẳng mở bị chặn của $\mathbb{R}$, với $a < b$. Khi đó tồn tại một dãy số đơn điệu tăng $a = b_0 < b_1 < b_2 < \ldots$ với $\lim_{n\to\infty} b_n = b$, và ta có thể viết $]a, b[= \bigcup_{n=1}^{\infty}]b_{n-1}, b_n]$, từ đó suy ra $]a, b[\in \mathcal{B}$, bởi vì $]b_{n-1}, b_n] \in \mathcal{B}$ với mọi n. Trong trường

hợp $b = +\infty$ ta vẫn có thể làm hệt như trên để chứng minh rằng $]a, +\infty[\in \mathcal{B}$. Khi $a = -\infty$, thì tồn tại một dãy số đơn điệu giảm $b = b_0 > b_1 > b_2 > \ldots$ với $\lim_{n\to\infty} b_n = -\infty$, và ta có thể viết $]-\infty, b[=]b_1, b_0[\cup \bigcup_{n=1}^{\infty}]b_{n+1}, b_n]$, từ đó suy ra $]-\infty, b[\in \mathcal{B}$. Đối với một đoạn thẳng đóng $[a, b]$, ta có $]-\infty, a[\in \mathcal{B}$, $]b, +\infty[\in \mathcal{B}$, và $[a, b] = \mathbb{R} \setminus (]-\infty, a[\cup]b, +\infty[)$, từ đó suy ra $[a, b] \in \mathcal{B}$. Các khẳng định ngược lại (các tập đóng sinh ra sigma-đại số $\mathcal{B}$, và các tập mở cũng sinh ra sigma-đại số $\mathcal{B}$) nhường cho bạn đọc làm bài tập. □

Định nghĩa 2.4. Hàm phân phối *xác suất của phân bố xác suất P_X trên $\mathbb{R}$ của một biến ngẫu nhiên X là hàm $\mathcal{F}_X : \mathbb{R} \to [0, 1]$ cho bởi công thức*

$$\mathcal{F}_X(x) := P(X \leq x) = P_X(]-\infty, x]). \tag{2.4}$$

Tất nhiên, hàm phân phối được xác định duy nhất bởi phân bố xác suất. Điều ngược lại cũng đúng: Nếu ta biết hàm phân phối $\mathcal{F}_X$, thì ta có thể tính được xác suất P_X của các đoạn thẳng đóng và nửa mở của $\mathbb{R}$ qua các công thức sau

$$P_X(]a, b]) = \mathcal{F}_X(b) - \mathcal{F}_X(a), \tag{2.5}$$

$$P_X([a, b]) = \mathcal{F}_X(b) - \lim_{x\to a-} \mathcal{F}_X(x), \tag{2.6}$$

và từ đó tính được xác suất của các tập con khác của $\mathbb{R}$.

Định lý 2.2. *Hàm phân phối $\mathcal{F}_X$ của một phân bố xác suất tùy ý trên $\mathbb{R}$ thỏa mãn 4 tính chất sau:*

1) Đơn điệu không giảm: $\mathcal{F}_X(x) \geq \mathcal{F}_X(y)$ với mọi $x \geq y$,

2) Liên tục bên phải: $\lim_{\epsilon\to 0+} \mathcal{F}_X(x + \epsilon) = \mathcal{F}_X(x)$ với mọi x,

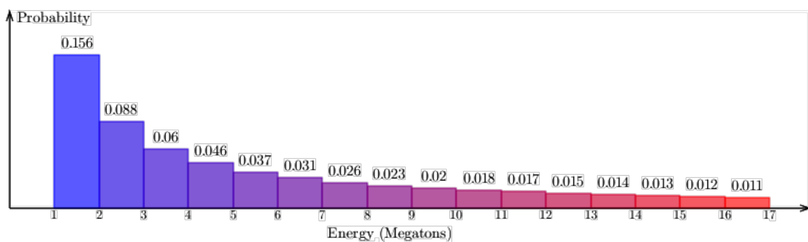
3) $\lim_{x\to -\infty} \mathcal{F}_X(x) = 0$,

4) $\lim_{x\to+\infty} \mathcal{F}_X(y) = 1.$
Ngược lại, mọi hàm số thực trên $\mathbb{R}$ *thỏa mãn 4 tính chất trên là hàm phân phối của một phân bố xác suất trên* $\mathbb{R}$

Chứng minh. Tính chất thứ nhất là hiển nhiên: nếu $x < y$ thì $\mathcal{F}_X(y) - \mathcal{F}_X(x) = P(x < X \leq y) \geq 0$. Tính chất thứ hai có thể phát biểu cách khác như sau: nếu $x_1 > x_2 > \ldots$ là một dãy số đơn điệu giảm với $x_n \to x$ khi n tiến tới vô cùng thì ta có $\lim_{n\to\infty} \mathcal{F}_X(x_n) = \mathcal{F}_X(x)$. Để thấy điều đó, ta có thể viết $\mathcal{F}_X(x_n) - \mathcal{F}_X(x) = P_X(]x, x_n]) = P_X(\bigcup_{k=n}^{\infty}]x_{k+1}, x_k]) = \sum_{k=n}^{\infty} P_X(]x_{k+1}, x_k])$. Chuỗi số dương

$$\sum_{k=1}^{\infty} P_X(]x_{k+1}, x_k])$$

là một chuỗi hội tụ, và bởi vậy *phần đuôi* $\sum_{k=n}^{\infty} P_X(]x_{k+1}, x_k])$ của nó tiến tới 0 khi n tiến tới vô cùng. Tính chất thứ 3 và tính chất thứ 4 có thể chứng minh một cách hoàn toàn tương tự. Khẳng định ngược lại là bài tập dành cho bạn đọc. □



Hình 2.1: Năng lượng của các thiên thạch đâm vào bầu khí quyển trái đất

Bài tập 2.1. Đồ thị 2.1 là biểu đồ phân bố xác suất (partial histogram, thiếu phần "đuôi") của mức năng lượng tỏa ra, tính theo đơn vị năng lượng megaton, của các thiên thạch lớn đâm vào bầu khí quyển của trái đất[1]. Hãy tính xác suất để một thiên thạch lớn đâm vào bầu khí quyển của trái đất có mức năng lượng tỏa ra không vượt quá 7 megaton.

2.1.4 Các loại phân bố xác suất trên $\mathbb{R}$

Trong nhiều công việc tính toán với biến ngẫu nhiên, ta có thể quên đi không gian xác suất ban đầu của biến ngẫu nhiên đó, mà chỉ cần biết đến phân bố xác suất trên $\mathbb{R}$ của nó. Các phân bố xác suất trên $\mathbb{R}$ có thể được chia làm 3 loại sau: rời rạc, liên tục, và hỗn hợp (nửa rời rạc nửa liên tục).

Định nghĩa 2.5. *Một phân bố xác suất P_X trên $\mathbb{R}$ được gọi là* **liên tục** *nếu như hàm phân phối xác suất $\mathcal{F}_X$ là hàm liên tục trên $\mathbb{R}$. Nó được gọi là* **liên tục tuyệt đối** *nếu như như tồn tại một hàm số $\rho_X : \mathbb{R} \to \mathbb{R}_+$ khả tích và không âm, sao cho với mọi $a \in \mathbb{R}$ ta có*

$$\mathcal{F}_X(a) = P_X(]-\infty, a]) = \int_{-\infty}^{a} \rho_X(x)dx$$

Hàm $\rho_X : \mathbb{R} \to \mathbb{R}_+$ thoả mãn điều kiện như trên gọi là **hàm mật độ** *của P_X.*

Ghi chú 2.1. Hàm mật độ của một phân bố xác suất liên tục tuyệt đối P_X trên $\mathbb{R}$ là duy nhất theo nghĩa xác suất: nếu P_X có hai hàm

[1] Số liệu của NASA năm 1994. Một thiên thạch lớn là một thiên thạch tỏa ra năng lượng ít nhất 1 megaton, bằng 1 quả bom hạt nhân nhỏ.

mật độ ρ_1 và ρ_2, thì $\rho_1 = \rho_2$ hầu khắp mọi nơi trên $\mathbb{R}$, tức là tập $\{x \in \mathbb{R}, \rho_1(x) \neq \rho_2(x)\}$ có độ đo Lebesgue bằng 0. Một phân bố xác suất có thể là liên tục mà không liên tục tuyệt đối. (Bài tập: xây dựng ví dụ). Tuy nhiên, trong thực tế, khi người ta nói đến một phân bố xác suất liên tục trên $\mathbb{R}$, thường được hiểu là nó liên tục tuyệt đối, tức là được cho bởi một hàm mật độ. Chú ý rằng hàm mật độ chính bằng đạo hàm của hàm phân phối xác suất (hầu khắp mọi nơi). Rất nhiều vấn đề trong thực tế có thể được mô hình hóa bằng các biến ngẫu nhiên với phân bố xác suất liên tục, ví dụ như nhiệt độ của nước biển, giá dầu hỏa, sản lượng điện, trọng lượng của trứng gà, v.v.

Định lý 2.3. *Giả sử X có phân bố xác suất liên tục với hàm mật độ ρ_X, và $f : \mathbb{R} \to \mathbb{R}$ là một đơn ánh khả vi liên tục trên $\mathbb{R}$ trừ một số hữu hạn các điểm. Khi đó $Y = f(X)$ cũng có phân bố xác suất liên tục, với hàm mật độ cho bởi công thức sau:*

$$\rho_Y(y) = \frac{\rho_X(x)}{|f'(x)|} \text{ tại điểm } y = f(x) \quad (2.7)$$

Công thức trên chẳng qua là công thức đổi biến trong tích phân, và sinh ra từ công thức $df(x) = f'(x)dx$.

Một điểm $x \in \mathbb{R}$ được gọi là một điểm **hạt** của một phân bố xác suất P_X nếu như $P_X(x) > 0$. Bổ đề sau cho thấy một phân bố xác suất là liên tục khi và chỉ khi nó không có điểm hạt:

Định lý 2.4. *Giả sử $\mathcal{F}_X$ là hàm phân phối xác suất của một phân bố xác suất P_X trên $\mathbb{R}$.*
i) Với mọi $x \in \mathbb{R}$ ta có

$$P_X(x) = \mathcal{F}_X(x) - \lim_{y \to x-} \mathcal{F}_X(y). \quad (2.8)$$

i) Hàm $\mathcal{F}_X$ là hàm liên tục trên $\mathbb{R}$ khi và chỉ khi $P_X(x) = 0$ với mọi $x \in \mathbb{R}$

Chứng minh. i) Nếu $x_0 < x_1 < x_2 < \ldots$ là một dãy số đơn điệu tăng có giới hạn là x, thì ta có

$$\begin{aligned}
&\lim_{n\to\infty} \mathcal{F}_X(x_n) = \mathcal{F}_X(x_0) + \lim_{n\to\infty} P_X(]x_0, x_n]) \\
&= \mathcal{F}_X(x_0) + \lim_{n\to\infty} \sum_{k=1}^{n} P_X(]x_{k-1}, x_k]) \\
&= \mathcal{F}_X(x_0) + \sum_{k=1}^{\infty} P_X(]x_{k-1}, x_k]) = \mathcal{F}_X(x_0) + P_X(\bigcup_{k=1}^{\infty}]x_{k-1}, x_k]) \\
&= \mathcal{F}_X(x_0) + P_X(]x_0, x[) = P_X(] - \infty, x[) \\
&= P_X(] - \infty, x]) - P_X(x) = \mathcal{F}_X(x) - P_X(x),
\end{aligned}$$

từ đó suy ra công thức trong bổ đề. Để chứng minh phần thứ hai của bồ đề trên, nhắc lại rằng hàm phân phối xác suất luôn luôn liên tục bên phải. Bởi vậy nó liên tục khi và chỉ khi nó liên tục bên trái, tức là khi và chỉ khi $P_X(x) = \mathcal{F}_X(x) - \lim_{y\to x-} \mathcal{F}_X(y) = 0$ với mọi $x \in \mathbb{R}$.

Trong trường hợp phân bố xác suất P_X không liên tục, gọi

$$K_X = \{x \in \mathbb{R} | P_X(x) > 0\} \tag{2.9}$$

là tập hợp các điểm hạt của nó (tức là tập hợp các điểm gián đoạn của hàm phân phối xác suất). Khi đó K_X là tập hữu hạn hoặc cùng lắm là đếm được, vì $P_X(A) = \sum_{x\in K_X} P_X(x) \leq 1$.

Định nghĩa 2.6. *Một phân bố xác suất P_X được gọi là* **rời rạc** *nếu như nó tập trung trên tập hợp các điểm hạt của nó: $P_X(A_X) = 1$, $P_X(\mathbb{R} \setminus A_X) = 0$.*

Ví dụ 2.6. Phân bố xác suất trên $\mathbb{R}$ của biến ngẫu nhiên "điểm kiểm tra" trong ví dụ "bài kiểm tra trắc nghiệm" ở mục trước là một phân

bố rời rạc tập trung ở 6 điểm: 0,1,2,3,4,5. (Bài tập: tính các xác suất của 6 điểm đó).

Giả sử P_X là một phân bố xác suất bất kỳ trên $\mathbb{R}$, với hàm phân phối $\mathcal{F}_X$. Khi đó ta có thể viết:

$$\mathcal{F}_X(x) = \mathcal{D}_X(x) + \mathcal{C}_X(x) \quad (2.10)$$

với $\mathcal{D}_X(x) = P_X(]-\infty, x] \cap K_X)$ gọi là **phần rời rạc** của $\mathcal{F}_X$, và $\mathcal{C}_X(x) = \mathcal{F}_X(x) - \mathcal{D}_X(x)$ gọi là **phần liên tục** của $\mathcal{F}_X$. Phân bố P_X được gọi là **hỗn hợp** nếu như cả hai phần rời rạc và liên tục đều khác 0. Nếu phần liên tục không phải là liên tục tuyệt đối (không viết được dưới dạng tích phân của một hàm không âm), thì ta có thể tách nó tiếp thành tổng của *phần liên tục tuyệt đối* và *phần liên tục kỳ dị*, nhưng chúng ta sẽ không đi vào chi tiết ở đây.

Ví dụ 2.7. Trong xe ô tô thường có kim chỉ mức xăng, dao động trong khoảng từ 0 (0%, tức là hết xăng) đến 1 (100%, bình xăng đầy). Mức xăng được kim chỉ vào có thể coi là một biến ngẫu nhiên nhận giá trị trong đoạn thẳng $[0, 1]$ với phân bố xác suất liên tục. Tuy nhiên, ở một số xe ô tô cũ, kim bị hỏng, có lúc nó chỉ đúng mức xăng nhưng có lúc nó bị tắc ở chỗ số 0 tuy rằng xe còn xăng. Khi đó, phân bố xác suất không còn là liên tục nữa mà là hỗn hợp, với "hạt" tại điểm 0.

Bài tập 2.2. Giả sử biến ngẫu nhiên X có phân bố xác suất liên tục với hàm mật độ ρ_X sau : $\rho_X(x) = 0$ khi $|x| > 1$ và $\rho_X(x) = 1 - |x|$ khi $|x| \leq 1$. Tìm hàm mật độ của phân bố xác suất của biến ngẫu nhiên $Y = \arcsin(x)$.

Bài tập 2.3. Giả sử biến ngẫu nhiên X có phân bố xác suất liên tục và đối xứng, theo nghĩa X và $-X$ có cùng phân bố xác suất. Chứng minh

rằng hàm phân phối xác suất của X thỏa mãn tính chất $\mathcal{F}_X(-x) + \mathcal{F}_X(x) = 1$ với mọi $x \in \mathbb{R}$. Điều này còn đúng không nếu phân bố xác suất của X không liên tục?

2.2 Một số phân bố xác suất thường gặp

Nhắc lại rằng, phân bố nhị thức với các tham số n, p là phân bố xác suất $P(k) = C_n^k p^k (1-p)^n$. trên không gian $\Omega = \{0, 1, \ldots, n\}$. Nó cũng có thể được coi như một phân bố rời rạc trên $\mathbb{R}$ tập trung tại các điểm $0, 1, \ldots, n$ với các xác suất như trên. Tương tự như vậy, phân bố Bernoulli với tham số p có thể được coi như một phân bố xác suất trên $\mathbb{R}$ tập trung tại hai điểm $0, 1$ (hoặc hai điểm nào đó khác), với các xác suất $P(1) = p$ và $P(0) = 1 - p$. Phân bố Bernoulli và phân bố nhị thức là những phân bố rất hay gặp trong thực tế. Ở đây, chúng ta sẽ thảo luận thêm một số phân bố rời rạc và liên tục phổ biến khác trên $\mathbb{R}$.

2.2.1 Phân bố hình học và phân bố nhị thức âm

Định nghĩa 2.7. Phân bố hình học *với tham số* p *(*$0 \leq p \leq 1$*) là phân bố xác suất rời rạc tập trung tại tập hợp các số tự nhiên, cho bởi công thức sau:*

$$P(k) = p(1-p)^{k-1} \ \ \forall\, k \in \mathbb{N}. \tag{2.11}$$

Ý nghĩa của phân bố hình học là: nó là phân bố xác suất của "số lần thử cho đến khi thành công", nếu như xác suất thành công của mỗi lần thử là p.

Ví dụ 2.8. Một người chơi trò tung vòng vào cổ chai, tung đến bao giờ trúng thì thôi. Xác suất để tung trúng mỗi lần là p. Gọi T là số lần phải tung cho đến khi tung trúng. Khi đó T là một biến ngẫu nhiên nhận giá trị trong $\mathbb{N}$. Xác suất để sao cho tung $k-1$ lần đầu trượt, nhưng lần thứ k trúng, là $(1-p)^{k-1}p$. Như vậy phân bố xác suất của T chính là phân bố hình học với tham số p.

Nếu thay vì tính số lần thử cho đến khi có 1 lần thành công, ta tính tổng số lần thử *thất bại* k cho đến khi có tổng cộng r lần thành công ($r \in \mathbb{N}$) thì ta có một biến ngẫu nhiên mới, nhận giá trị trong $\mathbb{Z}_+$, với phân bố xác suất sau:

$$P(k) = C^k_{k+r-1} p^r (1-p)^k$$

Nhị thức Newton C^k_{k+r-1} trong công thức trên là số cách chọn ra $r-1$ phần tử từ tập hợp $\{1, 2, \ldots, k+r-1\}$. (Mỗi cách chọn như vậy ứng với một tình huống, với k lần thất bại và $r-1$ lần thành công trong số $k+r-1$ lần thử đầu tiên, và lần thử thứ $k+r$ thành công). Các nhị thức Newton C^k_{k+r-1} còn có thể viết dưới dạng $C^k_{k+r-1} = \frac{(k+r-1).....(r+1).r}{k!} = (-1)^k \frac{(-r).(-r-1).....(-r-k+1)}{k!} = (-1)^k C^k_{-r}$, và chúng xuất hiện trong khai triển Taylor sau:

$$(1-q)^{-r} = \sum_{k=0}^{\infty} (-1)^k C^k_{-r} q^k$$

Trong khai triển Taylor trên, nếu đặt $q = 1-p$ và nhân cả hai vế với p^r, thì ta được

$$1 = \sum_{k=0}^{\infty} (-1)^k C^k_{-r} p^r (1-p)^k = \sum_{k=0}^{\infty} P(k)$$

Chú ý rằng khai triển Taylor trên có giá trị (và hội tụ khi $|q| < 1$) cả khi mà $r > 0$ không phải là số nguyên. Các công thức trên dẫn đến định nghĩa sau:

Định nghĩa 2.8. *Giả sử $0 < p < 1$ và $r > 0$. Khi đó phân bố xác suất rời rạc cho bởi công thức*

$$P(k) = C^k_{k+r-1} p^r (1-p)^k = (-1)^k C^k_{-r} p^r (1-p)^k \tag{2.12}$$

với mọi $k \in \mathbb{Z}_+$ được gọi là **phân bố nhị thức âm** *với các tham số r và p.*

Tất nhiên, phân bố hình học có thể coi là trường hợp đặc biệt của phân bố nhị thức âm, với $r = 1$ (và trên $\mathbb{N}$ thay vì trên $\mathbb{Z}_+$, tức là có cộng thêm 1 vào biến ngẫu nhiên).

Bài tập 2.4. Kiểm tra công thức sau: hàm phân phối xác suất của phân bố hình học với tham số p cho bởi công thức $\mathcal{F}(x) = 0$ nếu $x < 0$ và $\mathcal{F}(x) = 1 - (1-p)^{[x]}$ nếu $x \geq 0$. Ở đây $[x]$ là phần nguyên của số x.

2.2.2 Phân bố Poisson

Định nghĩa 2.9. *Một biến ngẫu nhiên X được gọi là có* **phân bố Poisson** *(đọc là Poa-Sông) với tham số λ, nếu như các giá trị của nó là các số nguyên không âm, và với mọi $k \in \mathbb{Z}_+$ ta có:*

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}. \tag{2.13}$$

Ghi chú 2.2. Phân bố Poisson mang tên của nhà toán học và vật lý người Pháp Siméon Denis Poisson (1781–1840). Trong lý thuyết xác

suất, Poisson được biết đến nhiều nhất bởi phân bố Poisson, và *quá trình Poisson* (một quá trình ngẫu nhiên ứng với phân bố này). Tên gọi *luật số lớn* (của các luật số lớn, mà chúng ta sẽ tìm hiểu trong Chương 4) cũng là do Poisson đặt ra.



Hình 2.2: Siméon Denis Poisson

Phân bố Poisson là giới hạn của phân bố nhị thức với các tham số

$p = \lambda/n$ và n, khi n tiến tới vô cùng. Thật vậy, ta có

$$\begin{aligned} C_n^k(\lambda/n)^k(1-\lambda/n)^{n-k} &= \frac{n!}{k!(n-k)!}(\lambda/n)^k(1-\lambda/n)^{n-k} \\ &= \frac{\lambda^k}{k!}\frac{n(n-1)\dots(n-k+1)}{n^k}(1-\lambda/n)^{-k}(1-\lambda/n)^n. \end{aligned}$$

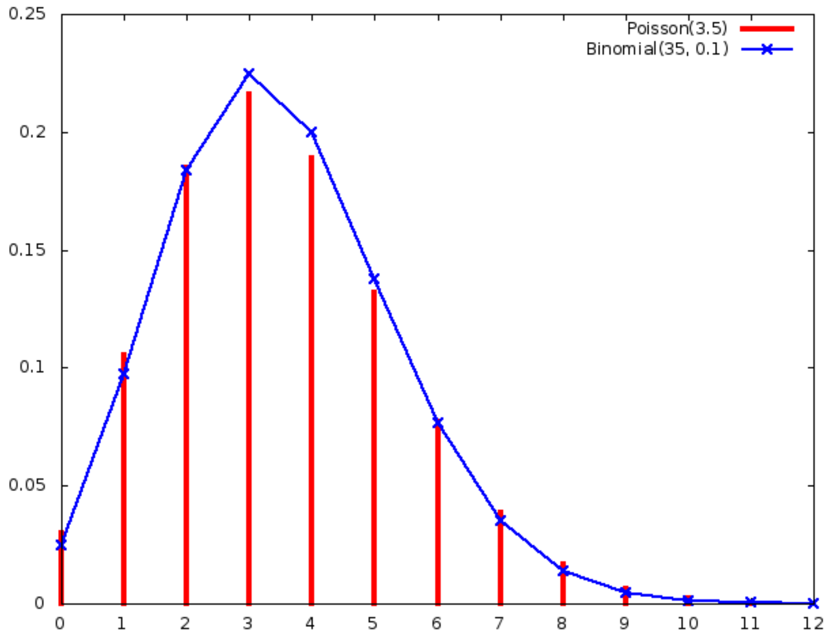
Khi n tiến tới vô cùng thì $(n(n-1)\dots(n-k+1)/n^k)(1-\lambda/n)^{-k}$ tiến tới 1 (k ở đây là cố định) và $(1-\lambda/n)^n$ tiến tới $e^{-\lambda}$, bởi vậy ta có

$$\lim_{n\to\infty} C_n^k.(\lambda/n)^k(1-\lambda/n)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}. \quad (2.14)$$

Xem đồ thị minh họa trên hình 2.3 cho trường hợp $\lambda = 3,5$, $n = 35$, $p = 0,1$.

Mô hình phân bố Poisson là mô hình thường được dùng cho các biến ngẫu nhiên dạng "số sự kiện xảy ra trong một khoảng thời gian nào đó".

Ví dụ 2.9. Biến ngẫu nhiên "số vụ tai nạn giao thông xảy ra trong một ngày" ở một vùng nào đó có thể được mô hình hóa bằng phân bố Poisson. Ta sẽ giả sử các tai nạn giao thông xảy ra một cách ngẫu nhiên, độc lập với nhau, và trung bình mỗi ngày có λ vụ tai nạn. Ta sẽ chia 24 tiếng đồng hồ trong ngày thành n khoảng thời gian (n là một số rất lớn), để sao cho có thể coi rằng trong mỗi khoảng thời gian có nhiều nhất 1 vụ giao thông xảy ra, và khả năng xảy ra tai nạn giao thông trong mỗi khoảng thời gian bằng λ/n. Khi đó tổng số tai nạn xảy ra trong ngày tuân theo phân bố nhị thức với các tham số $n, p = \lambda/n$, và khi cho n tiến tới vô cùng ta được phân bố Poisson. Tất nhiên phân bố Poisson không thể là phân bố xác suất chính xác của vấn đề (vì số người là hữu hạn, và số tai nạn bị chặn trên bởi



Hình 2.3: Các phân bố Poisson(3.5) và Binomial(35,0.1)

số người chứ không lớn tuỳ ý được), nhưng nó là phân bố gần đúng thuận tiện cho việc tính toán.

2.2.3 Phân bố đều (trường hợp liên tục)

Định nghĩa 2.10. *Giả sử a và b là hai số thực, với $b > a$. Khi đó* **phân bố đều** *(uniform distribution) trên đoạn thẳng $]a, b[$ là phân bố xác*

suất liên tục với hàm mật độ $\rho(x)$ sau:

$$\rho(x) = \begin{cases} \frac{1}{b-a} & \textit{khi } a \leq x \leq b \\ 0 & \textit{khi } x < a \textit{ hoặc } x > b \end{cases}. \tag{2.15}$$

Phân bố xác suất đều trên đoạn thẳng $]a, b[$ hay được ký hiệu là $U(a, b)$.

Ghi chú 2.3. Trong định nghĩa trên, thay vì lấy đoạn thẳng mở $]a, b[$, có thể lấy đoạn thẳng đóng $[a, b]$ hoặc đoạn thẳng nửa mở $]a, b]$ hoặc $[a, b[$ cũng được. Về mặt xác suất không có gì thay đổi.

Ví dụ 2.10. Vị trí của một người đi bộ trên một đoạn đường có thể được mô hình hóa bằng một biến ngẫu nhiên với phân bố đều, nếu như ta không có thông tin gì ngoài thông tin người đi bộ đang ở trên đoạn đường đó.

Khái niệm phân bố đều có thể mở rộng lên trường hợp nhiều chiều: không gian xác suất là một miền trong $\mathbb{R}^n$ ($n \geq 2$), và xác suất của một miền con tỷ lệ thuận với thể tích (n chiều) của miền con đó.

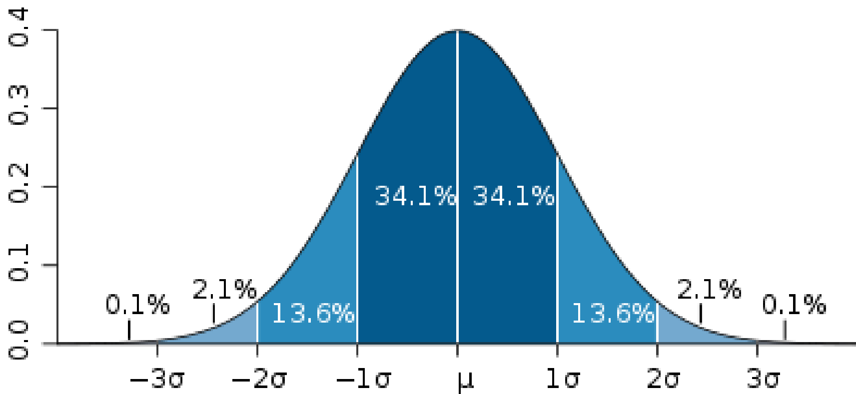
Bài tập 2.5. Giả sử X có phân bố đều $U(0, 1)$, và Y là một biến ngẫu nhiên bất kỳ. Chứng minh rằng tồn tại một hàm số g sao cho $g(X)$ và Y có cùng phân bố xác suất. (Bài tập này có ý nghĩa thực tế trong việc làm giả lập (simulation) các biến ngẫu nhiên: dùng random number generator (chương trình tạo số ngẫu nhiên) trên máy tính để giả lập một biến ngẫu nhiên với phân bố đều $U(0, 1)$, rồi qua đó giả lập được mọi phân bố xác suất, qua các hàm số thích ứng).

2.2.4 Phân bố normal

Định nghĩa 2.11. *Phân bố xác suất* **normal** *(còn gọi là phân bố* **chuẩn** *, hay phân bố* **Gauss***) trên $\mathbb{R}$ với trung điểm μ và độ lệch chuẩn σ là phân bố liên tục với hàm mật độ sau:*

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}). \tag{2.16}$$

Ký hiệu thường dùng để chỉ phân phối xác suất normal là: $\mathcal{N}(\mu, \sigma^2)$. Phân bố normal $\mathcal{N}(0,1)$ (với $\mu = 0$, $\sigma^2 = 1$) được gọi là **phân bố normal chuẩn tắc** .



Hình 2.4: Hàm mật độ của phân bố normal

Đồ thị của hàm mật độ của phân bố normal có hình cái chuông, và bởi vậy phân bố normal còn được gọi một cách nôm na là **phân bố hình cái chuông**. Trung điểm của cái chuông này chính là điểm $x = \mu$, và độ cao của chuông chính bằng $\frac{1}{\sigma\sqrt{2\pi}}$. Nếu σ càng nhỏ thì

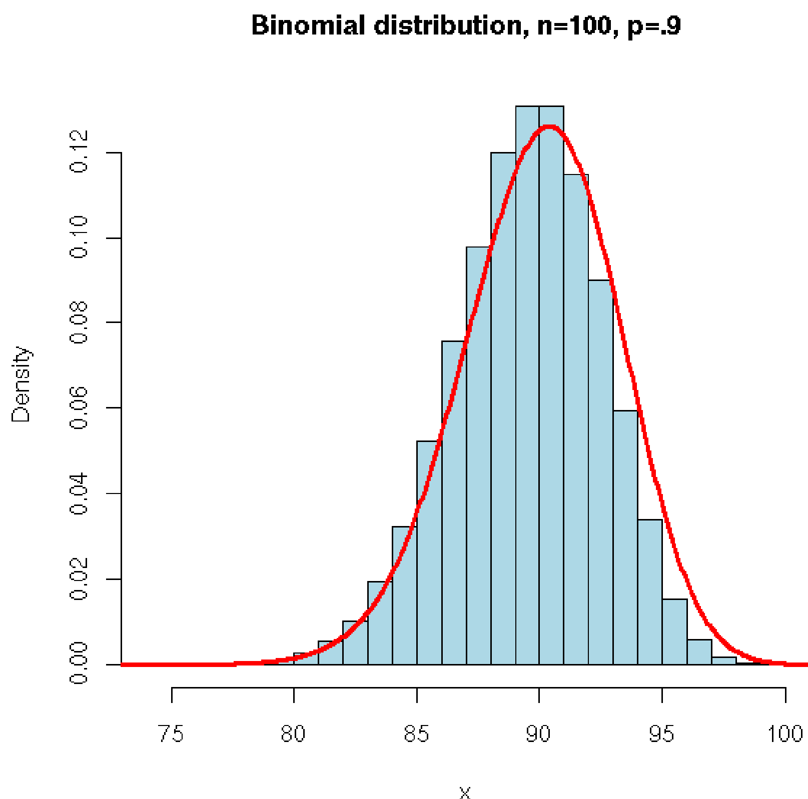
chuông càng cao và càng "hẹp", và ngược lại σ càng lớn thì chuông càng thấp và càng bè ra.

Hình vẽ minh họa 2.4 cho thấy hầu hết xác suất của một phân bố normal nằm trong đoạn $[\mu - 3\sigma, \mu + 3\sigma]$. Chỉ có không đến 0,3% nằm ngoài đoạn đó. Nói cách khác, nếu X là một biến ngẫu nhiên có phân bố xác suất normal với các tham số μ, σ, thì với xác suất 99,7% ta có thể tin rằng giá trị của X nằm trong đoạn $[\mu - 3\sigma, \mu + 3\sigma]$: $P(\mu - 3\sigma < X < \mu + 3\sigma) = 99,7\%$.

Phân bố normal là một trong những phân bố xác suất quan trọng nhất, vì nhiều phân bố xác suất gặp trong thực tế có dáng điệu khá giống phân bố normal, ví dụ như phân bố của chiều cao của đàn ông, phân bố của chỉ số IQ (chỉ số trí tuệ), phân bố của giá chứng khoán trong tương lai, v.v. Khi n tiến tới vô cùng và p cố định, thì dáng điệu của phân bố nhị thức với các tham số n, p cũng ngày càng gần giống phân bố normal. Ví dụ, lấy $p = 0,9$. Khi n nhỏ thì phân bố nhị thức với các tham số n và $p = 0,9$ có dáng điệu khác xa phân bố normal, nhưng khi $n = 100$, thì dáng điệu của phân bố nhị thức trông đã rất gần giống phân bố normal, như thể hiện trên Hình 2.5.

Các định lý giới hạn trung tâm mà chúng ta sẽ đề cập đến trong Chương 4 sẽ cho chúng ta cơ sở lý thuyết để hiểu tại sao có nhiều phân bố xác xuất trong thực tế trông giống phân bố normal.

Ví dụ 2.11. Hình 2.6 là **biểu đồ tần số** (histogram) của huyết áp của người, trong một thí nghiệm đo huyết áp 1000 người. **Tần số** (frequency) của một giá trị tức là số lần xuất hiện giá trị đó trong dãy số các kết quả. Nếu chúng ta coi không gian xác suất ở đây là có 1000 phần tử, với xác suất của một phần tử là 1/1000, thì bảng

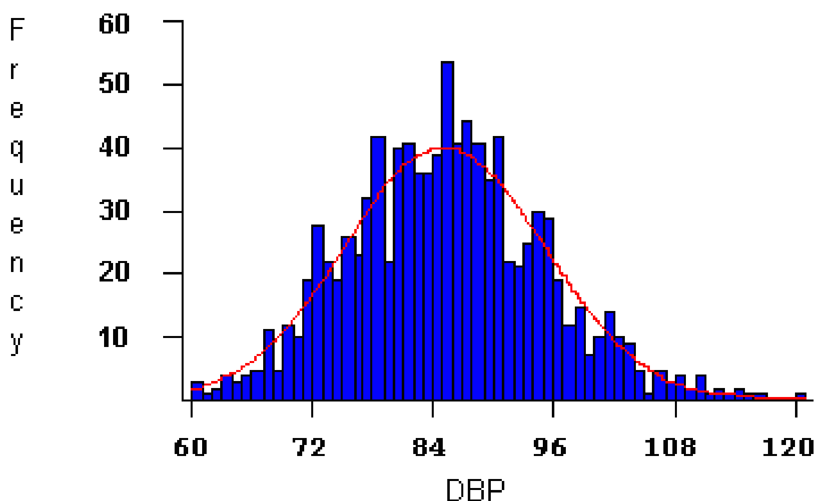


Hình 2.5: Phân bố nhị thức với $n = 100$, $p = 0,9$

tần số trên cho ta bảng phân bổ xác suất rời rạc của biến ngẫu nhiên "huyết áp" H: xác suất của sự kiện $H = x$ bằng **tần suất** (relative frequency[2]) của x. Tần suất là tần số chia cho tổng số (tức là chia

[2]Từ *frequency* tiếng Anh vừa có nghĩa là tần suất vừa có nghĩa tần số. Để phân

cho 1000 ở đây). Vì đồ thị có hình gần giống hình cái chuông, nên ta thấy phân bố xác suất của biến "huyết áp" trong thí nghiệm này có thể được xấp xỉ khá tốt bằng một phân bố normal.



Hình 2.6: Biểu đồ tần số huyết áp

Ghi chú 2.4. Để có một phân bố xác suất gần giống phân bố normal, cần phải có một sự "thuần nhất" nào đó trong biến ngẫu nhiên. Ví dụ, nếu ta có 1 thùng táo chín cùng một giống táo, thì khi xét biến ngẫu nhiên "đường kính của quả táo" trên thùng táo đó, ta có thể được một phân bố gần giống phân bố normal. Nhưng nếu ta trộn 2 thùng táo thuộc 2 giống táo khác nhau, một giống táo to một giống táo nhỏ, thì phân bố xác suất của biến "đường kính" trong đống táo

biệt, tần suất có khi được gọi là relative frequency, hoặc là frequency rate.

trộn lẫn này không còn là normal được nữa, mà nó phải có 2 "đỉnh", 1 đỉnh ứng với đường kính trung bình của giống táo to và 1 đỉnh ứng với đường kính trung bình của giống táo nhỏ.

Bài tập 2.6. Giả sử X là một biến ngẫu nhiên tuân theo luật phân bố normal $\mathcal{N}(\mu, \sigma^2)$. Chứng minh rằng biến ngẫu nhiên $Z = (X - \mu)/\sigma$ tuân theo phân bố normal chuẩn tắc $\mathcal{N}(0, 1)$.

2.2.5 Phân bố mũ

Định nghĩa 2.12. Phân bố mũ *(exponential distribution) với tham số λ là phân bố xác suất liên tục tuyệt đối trên $\mathbb{R}$ cho bởi hàm mật độ sau:*

$$\rho(x) = \begin{cases} \lambda e^{-\lambda x} & \textit{khi } x > 0 \\ 0 & \textit{khi } x \leq 0 \end{cases}. \tag{2.17}$$

Hàm phân bố xác suất $\mathcal{F}$ của phân bố này như sau: $\mathcal{F}(x) = 0$ khi $x \leq 0$, và khi $x > 0$ thì

$$\mathcal{F}(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}. \tag{2.18}$$

Phân bố mũ có thể được xem như là *dạng liên tục* của phân bố hình học: phân bố hình học là rời rạc còn phân bố mũ là liên tục, nhưng hàm phân phối xác suất của hai phân bố này có dáng điệu tương tự nhau.

Phân bố mũ có thể được dùng để làm mô hình xác suất cho những biến ngẫu nhiên kiểu "khoảng cách giữa hai lần xuất hiện", ví dụ như: khoảng cách thời gian giữa hai cú điện thoại gọi đến, khoảng cách giữa hai gen đột biến kế tiếp trên một dải DNA, v.v.

Bài tập 2.7. Giả sử biến ngẫu nhiên X có phân bố mũ với tham số λ, và $c > 0$. Chứng minh rằng cX cũng có phân bố mũ với tham số λ/c.

Bài tập 2.8. Giả sử biến ngẫu nhiên X có phân bố mũ với tham số λ, và s và t là hai số dương. Chứng minh rằng

$$P(X > s + t | X > s) = P(X > t)$$

Giải thích tại sao đẳng thức này gọi là tính chất *không có trí nhớ* (lack of memory property) của phân bố mũ.

Bài tập 2.9. Giả sử X là một biến ngẫu nhiên liên tục với hàm phân phối xác suất liên tục $f = \mathcal{F}_X$. Chứng minh rằng:
i) $f(X)$ có phân bố xác suất đều trên đoạn thẳng $[0, 1]$.
ii) $-\ln f(X)$ có phân bố mũ.

2.2.6 Phân bố Pareto

Vilfredo Pareto (1848–1923) là một nhà kinh tế người Italia. Ông ta quan sát thấy rằng, phân bố tài sản trên thế giới rất không đều, và "80% tài sản là do 20% người làm chủ" (80% nhân dân còn lại chỉ làm chủ 20% tài sản). Quan sát này mang tên *nguyên tắc Pareto* hay *nguyên tắc 80-20* (có khi nó còn trở thành *nguyên tắc 90-10*). Pareto đưa ra mô hình phân bố xác suất liên tục hóa sau cho biến ngẫu nhiên "giá trị tài sản của một người":

Định nghĩa 2.13. Phân bố Pareto *với tham số $\alpha > 0$ là phân bố liên tục trên $\mathbb{R}$ với hàm mật độ sau:*

$$\rho(x) = \begin{cases} \dfrac{\alpha}{x^{\alpha+1}} & \textit{khi } x \geq 1 \\ 0 & \textit{khi } x < 1 \end{cases} . \tag{2.19}$$



Hình 2.7: Vilfredo Pareto

Phân bố Pareto còn được dùng làm mô hình phân bố xác suất gần đúng cho rất nhiều biến ngẫu nhiên khác, vị dụ như: kích thước của các hạt cát, các thiên thạch, các khu dân cư, dự trữ dầu hỏa của các mỏ dầu, mức độ thiệt hại của các vụ tai nạn, v.v.

Bài tập 2.10. Chứng minh rằng nếu X có phân bố Pareto với tham số α, và $Y = X^s$ với $s > 0$, thì Y cũng có phân bố Pareto, và tìm tham số của phân bố này.

Bài tập 2.11. Giả sử X có phân bố xác suất đều $U(0,1)$. Chứng minh

rằng $Y = 1/(1 - X)$ có phân bố Pareto với tham số $\alpha = 1$.

2.3 Kỳ vọng của biến ngẫu nhiên

2.3.1 Trường hợp rời rạc

Khi ta có một biến ngẫu nhiên, ta có thể nghiên cứu các tính chất, đặc trưng của nó, để rút ra các thông tin, kết luận nào đó. Một trong những đặc trưng quan trọng nhất là giá trị kỳ vọng.

Định nghĩa 2.14. *Giá trị* **kỳ vọng** *của một biến ngẫu nhiên X, ký hiệu là $\mathbb{E}(X)$, chính là trung bình cộng của biến ngẫu nhiên đó trên không gian xác suất các tình huống.*

Từ định nghĩa có thể suy ra được rằng, hai biến ngẫu nhiên có cùng phân bố xác suất trên $\mathbb{R}$ thì có cùng kỳ vọng. Bởi vậy, thay vì nói về kỳ vọng của một biến ngẫu nhiên, ta cũng có thể nói về *kỳ vọng của một phân bố xác suất trên* $\mathbb{R}$.

Trong trường hợp không gian xác suất các tình huống là một tập hợp hữu hạn hoặc đếm được, $\Omega = \{\omega_1, \omega_2, \ldots\}$ với các xác suất $P(\omega_i)$ ($\sum_i P(\omega_i) = 1$), thì công thức tính giá trị kỳ vọng (trung bình cộng) của một biến ngẫu nhiên $X : \Omega \to \mathbb{R}$ là

$$\mathbb{E}(X) = \sum_i X(\omega_i) P(\omega_i). \tag{2.20}$$

Ví dụ 2.12. Trò chơi đề (một trò đánh bạc): trong 100 số đề sẽ chỉ có 1 số thắng, 99 số thua. Thắng thì được 70 lần tiền đặt cọc. Thua thì mất tiền đặt cọc. Nếu đặt cọc T tiền, thì kỳ vọng số tiền nhận lại được là

$99\% \times 0 + 1\% \times 70.T = 0,7.T$. Kỳ vọng lãi (lỗ) là $0,7.T - T = -0,3.T$. Tức là đặt cọc T tiền chơi đề, thì kỳ vọng là bị thua $0,3.T$.

Ví dụ 2.13. Giá trị kỳ vọng của phân bố Poisson $P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$ là $\mathbb{E}(X) = \lambda$. Thật vậy, $\mathbb{E}(X) = \sum_k kP(X = k) = \sum_k ke^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k\geq 1}\frac{\lambda^k}{(k-1)!} = e^{-\lambda}\lambda\sum_{k\geq 1}\frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\lambda e^{\lambda} = \lambda$.

Ví dụ 2.14. Giá trị kỳ vọng của phân bố hình học $P(T = k) = p(1 - p)^{k-1}$ là

$$\mathbb{E}(T) = \sum_{k=1}^{\infty} k.p.(1-p)^{k-1} = 1/p.$$

Điều này phù hợp với suy luận trực giác rằng, nếu xác suất để ném vòng một lần trúng cổ chai là p, thì trung bình phải ném vòng $1/p$ lần mới trúng cổ chai.

Ghi chú 2.5. Trong trường hợp không gian xác suất rời rạc $\Omega = \{\omega_1, \omega_2, \ldots\}$ có vô hạn các sự kiện, khi định nghĩa kỳ vọng, chúng ta đòi hỏi chuỗi $\sum_{i=1}^{\infty} X(\omega_i).P(\omega_i)$ phải là chuỗi *hội tụ tuyệt đối*, có nghĩa là chuỗi $\sum_{i=1}^{\infty} |X(\omega_i)|.P(\omega_i)$ phải hội tụ. Trong trường hợp chuỗi $\sum_{i=1}^{\infty} X(\omega_i).P(\omega_i)$ không hội tụ tuyệt đối, thì kỳ vọng không được xác định hoặc là bằng vô cùng. Lý do để đòi hỏi điều kiện hội tụ tuyệt đối là, chúng ta muốn tổng của chuỗi $\sum_{i=1}^{\infty} X(\omega_i).P(\omega_i)$ phải hữu hạn và không phụ thuộc vào thứ tự của các số trong tổng, tức là nếu có thay đổi cách đánh số các sự kiện, thì vẫn phải ra cùng một tổng. Các chuỗi thỏa mãn điều kiện này chính là các chuỗi hội tụ tuyệt đối.

Định lý 2.5. *Một số tính chất cơ bản của giá trị kỳ vọng:*
i) Kỳ vọng của một hằnh số c (biến ngẫu nhiên chỉ nhận 1 giá trị) chính

là hằng số đó:

$$\mathbb{E}(c) = c. \tag{2.21}$$

ii) Tuyến tính: Nếu X, Y là hai biến ngẫu nhiên và a, b là hai hằng số thì

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y). \tag{2.22}$$

iii) Đơn điệu: Nếu $X \geq 0$ thì $\mathbb{E}(X) \geq 0$. Tổng quát hơn,

$$X \geq Y \Rightarrow \mathbb{E}(X) \geq \mathbb{E}(Y). \tag{2.23}$$

Định lý trên đúng trong trường hợp tổng quát, khi mà các giá trị kỳ vọng được xác định. Chứng minh của nó trong trường hợp rời rạc tương đối hiển nhiên.

Khi chúng ta sử dụng hai mô hình không gian xác suất khác nhau để nghiên cứu cùng một biến ngẫu nhiên, thì không phải vì thế mà kỳ vọng của nó thay đổi. Nói một cách chính xác hơn, ta có:

Định lý 2.6. *Giả sử $X : (\Omega, P) \to \mathbb{R}$ là một biến ngẫu nhiên với không gian xác suất Ω, và $\phi : (\Omega_1, P_1) \to (\Omega, P)$ là một ánh xạ bảo toàn xác suất từ một không gian xác suất (Ω_1, P_1) lên (Ω_1, P_1). Đặt $X_1 = X \circ \phi : (\Omega_1, P_1) \to \mathbb{R}$ là biến ngẫu nhiên giống X nhưng với không gian xác suất (Ω_1, P_1). Khi đó*

$$\mathbb{E}(X_1) = \mathbb{E}(X). \tag{2.24}$$

Định lý trên cũng đúng trong trường hợp tổng quát. Chứng minh của nó tương đối hiển nhiên trong trường hợp Ω và Ω_1 là các không gian xác suất rời rạc, và là bài tập dành cho bạn đọc.

Bài tập 2.12. Một doanh nghiệp đầu tư phát triển một sản phẩm mới, xác suất thành công là 30%. Chi phí đầu tư bỏ ra là 100 nghìn USD. Nếu không thành công thì mất chi phí đầu tư mà không thu về được gì, nhưng nếu thành công thì thu về được 1 triệu (trước khi trừ đi chi phí đầu tư). Tính kỳ vọng lợi nhuận từ vụ đầu tư này.

Bài tập 2.13. Xây dựng một ví dụ đơn giản với hai biến ngẫu nhiên X, Y rời rạc sao cho $\mathbb{E}(XY) \neq \mathbb{E}(X)\mathbb{E}(Y)$.

Bài tập 2.14. Trong một rổ có 99 quả bóng đánh số từ 1 đến 99. Lôi ra từ trong rổ 5 quả bóng một cách ngẫu nhiên. Gọi X là số nhỏ nhất hiện lên trên 5 quả bóng được lôi ra, và Y là số lớn nhất hiện lên.
i) Tính phân bố xác suất của các biến ngẫu nhiên X và Y.
ii) Chứng minh rằng, với mọi $m, n \in \mathbb{N}, m \leq n$, ta có $\sum_{k=m}^{n} C_k^m = C_{n+1}^{m+1}$.
iii) Dùng ii) để tính $\mathbb{E}(X)$.

Bài tập 2.15. Một người tập bóng rổ, đứng từ một chỗ ném bóng vào rổ 6 lần. Xác suất ném trúng mỗi lần là 2/3. Gọi X là số lần ném trúng, Y là số lần ném trượt, và $Z = X - Y$. Hãy tính kỳ vọng $\mathbb{E}(Z)$ của Z bằng hai cách khác nhau: một cách thông qua phân bố xác suất của Z, và một cách không dùng đến phân bố xác suất của Z.

Bài tập 2.16. (Entropy). Giả sử có 1 trò chơi giữa hai người A và B như sau: A chọn 1 số tự nhiên trong các số từ 1 đến 2^n (n là một số cố định nào đó), và B phải tìm xem là số nào. B có thể hỏi A bất cứ câu hỏi nào về số mà A chọn, mà có thể phát biểu lại dưới dạng "số đó có thuộc một tập con X nào đó của tập các số tự nhiên trên hay không, và A sẽ trả lời "có" hoặc "không" cho các câu hỏi của B. (Ví dụ có thể hỏi: số đó có lớn hơn 5 hay không, là số chẵn hay không, v.v.)

i) Chỉ ra một chiến thuật (mộ cách hỏi), để sau khi hỏi đúng n lần, B tìm được số mà A chọn. (Số n ở đây được gọi là *entropy*, hay là *lượng thông tin*).
ii) Chứng minh rằng, với bất kỳ chiến thuật nào của B, thì kỳ vọng về số lần phải hỏi cho đến khi tìm được số mà A chọn là một số lớn hơn hoặc bằng n.
(Đầu tiên hãy thử làm cho các trường hợp $n = 2$, $n = 3$, rồi làm cho trường hợp tổng quát).

2.3.2 Trường hợp tổng quát: tích phân trên không gian xác suất

Trong trường hợp tổng quát, công thức tính giá trị kỳ vọng được viết dưới dạng **tích phân Lebesgue** của X trên không gian xác suất (Ω, P):

$$\mathbb{E}(X) = \int_{\Omega} X dP. \tag{2.25}$$

Định nghĩa của tích phân Lebesgue như sau. Giả sử có một hàm số $F : (\Omega, P) \to \mathbb{R}$ đo được trên một không gian xác suất (Ω, P) với độ đo xác suất P. Nhắc lại rằng, tính chất đo được có nghĩa là tồn tại $P(F^{-1}(]a, b]))$ với mọi $a, b \in \mathbb{R}, a < b$.

Trước hết ta xét trường hợp F là một hàm bị chặn: tồn tại một số dương $M \in \mathbb{R}_+$ sao cho $|F(\omega)| < M$ với mọi $\omega \in \Omega$.

Một phân hoạch (sự chia nhỏ) của đoạn thẳng $]-M, M]$ là một dãy số $a_0 = -M < a_1 < a_2 < \ldots < a_n = M$ hữu hạn đơn điệu tăng nào đó, sao cho số đầu bằng $-M$ và số cuối bằng M. Nói cách khác, ta chia đoạn thẳng $]-M, M]$ thành một hợp không giao nhau của

các đoạn thẳng nửa mở $]a_i, a_{i+1}]$. Khi có một phân hoạch như vậy, ký hiệu là σ, ta có thể lập hai số sau:

$$I_\sigma(F) = \sum_{i=0}^{n} a_i.P(F^{-1}(]a_i, a_{i+1}])), \tag{2.26}$$

và

$$J_\sigma(F) = \sum_{i=0}^{n} a_{i+1}.P(F^{-1}(]a_i, a_{i+1}])). \tag{2.27}$$

Ký hiệu Σ là tập hợp tất cả các phân hoạch của đoạn thẳng $]-M, M]$. Dễ thấy rằng

$$I_\sigma(g) \leq J_\delta(F) \ \ \forall\, \sigma, \delta \in \Sigma.$$

(Bài tập: Chứng minh bất đẳng thức trên). Hơn nữa, nếu phân hoạch σ thỏa mãn tính chất $a_{i+1} - a_i < \epsilon$ với mọi i, thì ta cũng có $J_\sigma(g) - I_\sigma(g) < \epsilon$. Từ đó suy ra $\sup_{\sigma\in\Sigma} I_\sigma(F) = \inf_{\delta\in\Sigma} J_\delta(F)$. Theo định nghĩa, tích phân Lebesgue của F trên (Ω, P) chính là giá trị chung đó:

$$\int_\Omega F dP = \sup_{\sigma\in\Sigma} I_\sigma(F) = \inf_{\delta\in\Sigma} J_\delta(F). \tag{2.28}$$

Trong trường hợp F không bị chặn, thì đầu tiên ta thay F bằng các hàm bị chặn

$$F_{M,N}(\omega) := \min(\max(-N, F(\omega)), M), \tag{2.29}$$

$(M, N > 0)$, rồi định nghĩa

$$\int_\Omega F dP = \lim_{M,N\to+\infty} \int_\Omega F_{M,N} dP \tag{2.30}$$

nếu như giới hạn đó tồn tại. Trong trường hợp giới hạn đó tồn tại và hữu hạn, thì ta nói F là hàm khả tích. **Khả tích** có nghĩa là định nghĩa được tích phân, và các cách định nghĩa khác nhau (qua các cách lấy giới hạn khác nhau) cho cùng một kết quả hữu hạn. Hàm F khả tích khi và chỉ khi giá trị tuyệt đối của nó có tích phân hữu hạn: $\int_\Omega |F| dP < \infty$. (Đây là một định lý trong giải tích, chứng minh không khó).

Trong trường hợp Ω là một miền trong $\mathbb{R}^n$ với thể tích bằng 1, phân bố xác suất P là phân bố đều trên đó (xác suất của một miền con của Ω là thể tích của miền con đó), và F là một hàm liên tục bị chặn, thì tích phân Lebesgue trùng với tích phân (Riemann) nhiều chiều thông thường. Trong trường hợp tổng quát, thì tích phân Lebesgue là mở rộng của khái niệm tích phân Riemann.

Tất nhiên, trong trường hợp $\Omega = \{\omega_1, \omega_2, \ldots\}$ là một không gian xác suất rời rạc, ta có

$$\int_\Omega F dP = \sum_i F(\omega_i).P(\omega_i), \tag{2.31}$$

và (nếu Ω có vô hạn phần tử) F khả tích khi và chỉ khi chuỗi

$$\sum_i F(\omega_i).P(\omega_i)$$

hội tụ tuyệt đối.

Tương tự như tích phân Riemann thông thường, tích phân Lebesgue trên không gian xác suất có tính chất đơn điệu, tuyến tính, và giao hoán với phép lấy giới hạn của một dãy hàm hội tụ đều:

Định lý 2.7. *Giả sử F, G và F_n là các hàm đo được trên một không gian xác suất (Ω, P).*

i) Nếu $F \geq 0$ (hầu khắp mọi nơi trên Ω) thì $\int_\Omega FdP \geq 0$. Tổng quát hơn, nếu $F \geq G$ thì $\int_\Omega FdP \geq \int_\Omega GdP$.
ii) Nếu F_n hội tụ đều đến F trên Ω, có nghĩa là $\lim_{n\to\infty} \sup_{\omega\in\Omega} |F_n(\omega) - F(\omega)| = 0$, thì $\lim_{n\to\infty} F_n dP = \int_\Omega FdP$.
iii) Với hai số thực a, b bất kỳ, ta có

$$\int_\Omega (aF + bG)dP = a\int_\Omega FdP + b\int_\Omega GdP. \tag{2.32}$$

Hai khẳng định đầu tiên của định lý trên suy ra trực tiếp từ định nghĩa của tích phân Lebesgue. Khẳng định thứ ba có thể kiểm tra trực tiếp dễ dàng trong trường hợp F và G chỉ nhận một số hữu hạn các giá trị. Trong trường hợp tổng quát, ta có thể xấp xỉ F và G bằng các hàm chỉ nhận một số hữu hạn các giá trị, sau đó lấy giới hạn. □

Định lý sau, gọi là *định lý hội tụ bị chặn Lebesgue* (Lebesgue dominated convergence theorem), là một định lý hay được sử dụng trong việc nghiên cứu các tích phân Lebesgue:

Định lý 2.8 (Lebesgue). *Giả sử $F_n : (\Omega, P) \to \mathbb{R}$ là một dãy hàm đo được trên không gian xác suất (Ω, P) thỏa mãn hai điều kiện sau:*
i) $|F_n| \leq G$ với mọi n, trong đó G là một hàm khả tích trên Ω.
ii) F_n hội tụ hầu khắp mọi nơi đến một hàm đo được F trên Ω, có nghĩa là tập các điểm $\omega \in \Omega$ sao cho $\lim_{n\to\infty} F_n(\omega) = F(\omega)$ có độ đo bằng 1.
Khi đó ta có

$$\int_\Omega FdP = \lim_{n\to\infty} \int_\Omega F_n dP. \tag{2.33}$$

Sơ lược chứng minh. Vì $|F_n| \leq G$ nên ta cũng có $|F| \leq G$ hầu khắp mọi nơi. Lấy một số $\delta > 0$ nhỏ tùy ý. Đặt $A_n = \{\omega \in \Omega | |F_n(\omega) -$

$F(\omega)| > \delta\}$. Ta có

$$\begin{aligned} |\int_\Omega FdP - \int_\Omega F_n dP| &\leq \int_\Omega |F - F_n| dP \\ &\leq \int_{\Omega \setminus A_n} \delta dP + \int_{A_n} 2G dP \leq \delta + 2\int_{A_n} G dP. \end{aligned}$$

Để chứng minh $|\int_\Omega FdP - \int_\Omega F_n dP|$ tiến tới 0 khi n tiến tới vô cùng, ta chỉ cần chứng minh $\int_{A_n} GdP$ tiến tới 0 khi n tiến tới vô cùng với mọi δ. Vì F_n hội tụ hầu khắp mọi nơi đến F trên Ω, nên tập hợp các điểm mà nằm trong vô số các tập A_n có độ đo bằng 0. Do đó $P(A_n)$ tiến tới 0 khi n tiến tới vô cùng (xem khẳng định thứ hai của Bài tập 1.22), từ đó suy ra $\int_{A_n} GdP$ tiến tới 0 khi n tiến tới vô cùng. □

Định lý 2.6 về sự bảo toàn giá trị kỳ vọng dưới ánh xạ bảo toàn xác suất có thể được phát biểu lại dưới dạng định lý về sự bảo toàn tích phân Lebesgue dưới dánh xạ bảo toàn xác suất:

Định lý 2.9. *Giả sử $F : (\Omega, P) \to \mathbb{R}$ là một hàm khả tích trên không gian xác suất (Ω, P), và $\phi : (\Omega_1, P_1) \to (\Omega, P)$ là một ánh xạ bảo toàn xác suất. Khi đó $F \circ \phi$ là hàm khả tích trên (Ω_1, P_1) và*

$$\int_{\Omega_1} (F \circ \phi) dP_1 = \int_\Omega FdP. \tag{2.34}$$

Chứng minh của định lý trên suy ra trực tiếp từ định nghĩa tích phân Lebesgue. □

2.3.3 Kỳ vọng của phân bố xác suất trên $\mathbb{R}$

Đôi khi, ta sẽ ký hiệu tích phân $\int_\Omega FdP$ thành $\int_{\omega\in\Omega} F(\omega)dP$, hoặc là $\int_\Omega F(\omega)dP(\omega)$, để chỉ rõ hơn về việc lấy tích phân theo biến nào.

Theo định nghĩa, **kỳ vọng của một phân bố xác suất** P_X trên $\mathbb{R}$ là $\int_{x\in\mathbb{R}} x dP_X$.

Định lý 2.10. *i) Kỳ vọng của một biến ngẫu nhiên X bằng kỳ vọng của phân bố xác suất P_X của biến ngẫu nhiên đó:*

$$\mathbb{E}(X) = \int_{x\in\mathbb{R}} x dP_X. \tag{2.35}$$

ii) Nếu P_X là một phân bố liên tục với hàm mật độ ρ_X, thì ta có:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\rho_X(x)dx. \tag{2.36}$$

iii) Nếu g là một hàm số thực thì

$$\mathbb{E}(g(X)) = \int_{x\in\mathbb{R}} g(x)dP_X = \int_{-\infty}^{\infty} g(x)\rho_X(x)dx. \tag{2.37}$$

Khẳng định đầu tiên của định lý trên chẳng qua là trường hợp đặc biệt của tính chất bảo toàn kỳ vọng qua ánh xạ bảo toàn xác suất. Thật vậy, ta có thể viết $X = Id \circ X$, trong đó Id là hàm đồng nhất trên $\mathbb{R}$: $Id(x) = x$. Do đó kỳ vọng của X bằng kỳ vọng của hàm Id trên $\mathbb{R}$ với phân bố xác suất P_X, và ta có công thức (2.35). Khẳng định thứ hai là hệ quả của khẳng định thứ nhất trong trường hợp liên tục tuyệt đối. Khẳng định thứ ba cũng suy ra từ tính chất bảo toàn kỳ vọng qua ánh xạ bảo toàn xác suất, tương tự như khẳng định thứ nhất. □

Ví dụ 2.15. Giá trị kỳ vọng của phân bố xác suất normal $\mathcal{N}(\mu, \sigma^2)$ bằng μ.

Ví dụ 2.16. Giả sử giá 1kg vàng vào thời điểm T là 35000 (USD). Tại thời điểm T, thì giá 1kg vàng cho thời điểm $T + 1$ chưa được

biết, và có thể coi là một biến ngẫu nhiên X. Giả sử rằng X có phân bố (gần như) normal với kỳ vọng 35000 và độ lệch chuẩn 400. Hỏi rằng, vào thời điểm T, giá trị của quyền mua 1kg với giá 35000 tại thời điểm $T+1$ là bao nhiêu? Quyền mua (call) vàng là một chứng khoán phái sinh, cho phép người sở hữu nó mua vàng với giá cố định trước, tại một thời điểm trong tương lai, nhưng không bắt buộc phải mua. Gọi giá trị của quyền mua này tại thời điểm $T+1$ là Y. Khi đó $Y = \max(0, X - 35000)$, tức là nếu giá vàng lúc đó dưới 35000 thì giá trị của quyền mua bằng 0, còn nếu giá vàng trên 35000 thì giá trị của quyền mua bằng sự chênh lệch giữa giá vàng và giá ghi trong quyền mua. Giá trị của quyền mua này tại thời điểm T được coi bằng kỳ vọng của Y. Như vậy giá trị này bằng

$$\begin{aligned}\mathbb{E}(Y) &= \frac{1}{400\sqrt{2\pi}} \int_{35000}^{\infty} (x - 35000).\exp\left(-\frac{(x-35000)^2}{2.400^2}\right) dx \\ &= \frac{1}{400\sqrt{2\pi}} \int_0^{\infty} x \exp\left(-\frac{x^2}{2.400^2}\right) dx = \frac{400}{\sqrt{2\pi}} \int_0^{\infty} \exp(-z) dz \\ &= \frac{400}{\sqrt{2\pi}} \approx 160.\end{aligned}$$

Bài tập 2.17. Giả sử Y là một biến ngẫu nhiên liên tục với hàm mật độ sau: $\rho_Y(x) = c \sin x$ khi $x \in]0, \pi[$, và $\rho_Y(x) = 0$ tại các điểm khác.
i) Hãy tính c.
ii) Hãy tính $\mathbb{E}(Y)$
iii) Thử nghĩ một vấn đề có thể xảy ra trong thực tế với phân bố xác suất này.

Bài tập 2.18. Tính kỳ vọng của phân bố Pareto (2.19) với tham số $\alpha > 1$. (Khi $\alpha \leq 1$ thì kỳ vọng bằng vô cùng).

2.3.4 Giá trị kỳ vọng hình học

Trong các tài liệu về xác suất ít khi nhắc tới kỳ vọng hình học. Nhưng khái niệm này cũng rất quan trọng, bởi vậy chúng ta sẽ đề cập nó ở đây. Giá trị kỳ vọng ứng với trung bình cộng, còn giá trì kỳ vọng hình học ứng với trung bình nhân. Một ví dụ đơn giản sau đây cho thấy sự quan trọng của trung bình nhân trong thực tế.

Ví dụ 2.17. Giả sử giá nhà dao động trong 4 năm như sau. Năm đầu tiên giảm 15%, năm thứ hai tăng 35%, năm thứ ba giảm 20%, năm thứ tư tăng 20%. Hỏi xem trong 4 năm đó giá nhà tăng lên (hay giảm đi) trùng bình mỗi năm bao nhiêu %? Nếu ta lấy trung bình cộng thì được (-15% + 35% - 20% + 20%)/ 4 = 5% một năm. Nhưng con số đó có phản ánh chính xác sự đi lên của giá nhà trong 4 năm không? Nếu gọi giá lúc đầu là X, thì sau năm đầu giá là (1-15%)X, sau năm thứ hai giá là (1+35%)(1-15%)X, sau năm thứ ba giá là (1-20%)(1+35%)(1-15%)X, sau 4 năm giá là (1+20%)(1-20%)(1+35%)(1-15%)X = 1,1016 X. Tức là sau 4 năm giá nhà chỉ tăng lên có 10,16%, chứ không phải 20% (= 4 lần 5%) như là người ta tưởng! Để có cái nhìn chính xác về mức độ tăng trưởng trung bình hàng năm trong giai đoạn 4 năm, cần phải lấy trung bình nhân của các con số 1+20%, 1-20%, 1+35%, 1-15% rồi trừ đi 1. Kết quả là 2,449% một năm.

Như chúng ta biết, nếu có một dãy các số dương $a_1, \ldots, a_n$, $a_i > 0$ với mọi i, thì ngoài giá trị trung bình cộng $(\sum a_i)/n$, chúng ta còn có thể nói đến trung bình nhân: $(\prod_i a_i)^{1/n}$. Từ tiếng Anh cho trung bình nhân là geometric mean, nếu dịch từng chữ ra tiếng Việt thì là

"trung bình hình học", còn trung bình cộng là "trung bình số học". Trung bình nhân có thể được định nghĩa qua trung bình cộng và qua hàm logarithm $\ln$, và hàm ngược của hàm $\ln$, tức là hàm $\exp$:

$$(\prod_i a_i)^{1/n} = \exp(\sum_i (\ln a_i)/n). \tag{2.38}$$

Hàm $\ln$ là hàm lõm trên nửa đường thẳng dương (đạo hàm bậc hai của nó bằng $-1/x^2$ là một hàm âm), bởi vậy ta có:

$$\frac{\sum_i \ln a_i}{n} \leq \ln\left(\frac{\sum_i a_i}{n}\right)$$

Lấy $\exp$ của hai vế của bất đẳng trên, ta được bất đẳng thức quen thuộc sau: Trung bình nhân luôn luôn nhỏ hơn hoặc bằng trung bình cộng:

$$\left(\prod_i a_i\right)^{1/n} \leq \frac{\sum_i a_i}{n}. \tag{2.39}$$

Dấu bằng xảy ra khi và chỉ khi tất cả các số a_i bằng nhau.

Nếu thay vì một dãy các số dương, ta có một biến ngẫu nhiên X mà các giá trị đều dương, thì ta cũng có thể làm tương tự như trên, và kết quả gọi là giá trị kỳ vọng hình học của X:

Định nghĩa 2.15. *Nếu X là một biết ngẫu nhiên chỉ nhận các giá trị dương, thì* **giá trị kỳ vọng hình học** *của X, ký hiệu là $\mathbb{G}(X)$, được cho bở công thức sau:*

$$\mathbb{G}(X) = \exp(\mathbb{E}(\ln X)) = \exp(\int_\Omega \ln(X)dP). \tag{2.40}$$

Định lý 2.11. *Giá trị kỳ vọng hình học luôn nhỏ hơn hoặc bằng giá trị kỳ vọng:*

$$\mathbb{G}(X) \leq \mathbb{E}(X). \tag{2.41}$$

Dấu bằng xảy ra khi và chỉ khi F là hằng số hầu khắp mọi nơi trên không gian xác suất, tức là tồn tại một số thực dương c sao cho $P(X = c) = 1$.

Định lý trên là trường hợp riêng của bất đẳng thức Jensen phát biểu như sau:

Định lý 2.12 (Bất đẳng thức Jensen). *Nếu f là một hàm lồi, và X là một biến ngẫu nhiên bất kỳ, thì*

$$\mathbb{E}(f(X)) \geq f(\mathbb{E}(X)). \quad (2.42)$$

Ví dụ 2.18. Giả sử có một cơ hội đầu tư như sau. Khả năng thắng/thua là 50%/50%, sau 1 tháng biết kết quả. Nếu thắng thì lãi 100%, nếu thua thì lỗ 50% tiền bỏ ra. (Trên thị trường chứng khoán có những trường hợp tương tự như vậy, ví dụ như 1 hãng công nghệ sinh học khi đang đợi kết quả thí nghiệm lâm sàng của một loại thuốc chữa ung thư, nếu thành công thì giá trị cổ phiếu của hãng có thể tăng hơn gấp đôi, nếu thất bại thì giá trị cũng có thể mất trên 50%). Hỏi đối với người đầu tư thì có nên đầu tư vào những cơ hội như vậy không, và nếu nên thì nên đầu tư với nhiều nhất nhiêu % vốn đầu tư (để đạt kỳ vọng lợi nhuận cao nhất, giả sử là không có các cơ hội đầu tư khác)?

Trước hết, ta có thể tính giá trị kỳ vọng của lợi nhuận của đầu tư theo cơ hội trên, với 1 đơn vị vốn bỏ ra. Gọi L là biến "lợi nhuận", ta có 2 khả năng: hoặc $L = 1$ hoặc $L = -1/2$, mỗi khả năng có xác suất 50%. Như vậy kỳ vọng lợi nhuận trên 1 đơn vị vốn bỏ ra là: $\mathbb{E}(L) = 50\%.1 + 50\%.(-1/2) = 0,25$ Kỳ vọng lợi nhận ở đây là dương

và khá lớn (bằng 25% vốn bỏ ra), nên đây là cơ hội nên đầu tư, trừ khi có những cơ hội khác tốt hơn. (Lãi 25% trong một tháng có thể gọi là siêu lợi nhuận).

Câu hỏi thứ hai là nhà đầu tư nên đầu tư vào đó nhiều nhất là bao nhiêu phần trăm vốn đầu tư? Nếu giả sử đầu tư toàn bộ 100% vốn. Khi đó có 2 khả năng, hoặc là tổng số vốn tăng lên gấp đôi, hoặc là giảm đi còn 1 nửa, với xác suất của mỗi khả năng là 50%. Nhưng nếu một nhà đầu tư mà làm như vậy 2 lần liên tiếp, 1 lần thắng một lần thua, thì sau hai lần số vốn lại về như cũ không tăng trưởng được gì cả. Muốn đảm bảo cho vốn tăng trưởng "về lâu về dài", cái cần tính đến không phải là giá trị kỳ vọng của vốn sau mỗi lần đầu tư, mà là giá trị kỳ vọng hình học. Nếu giả sử chỉ có 1 cơ hội đầu tư duy nhất như trên, thì giá trị kỳ vọng hình học của vốn có được sau khi đầu tư Y tiền vào đó trên tổng số X tiền sẽ là: $\sqrt{(X-Y/2)(X+Y)}$ Để tối ưu hóa giá trị kỳ vọng hình học tức là tìm Y sao cho $\sqrt{(X-Y/2)(X+Y)}$ đặt cực đại, với X cho trước. Kết quả là $Y = X/2$, và khi đó giá trị kỳ vọng hình học của vốn sau khi đầu tư là $\sqrt{(X-X/4)(X+X/2)} = 1,061.X$ Như vậy, kỳ vọng lợi nhuận của một cơ hội đầu tư như trên, tính trên toàn bộ vốn của nhà đầu tư, chỉ có không quá 6,1% chứ không phải 25%.

Định lý 2.13. *Giá trị kỳ vọng hình học có những tính chất sau:*
Tính đơn điệu: nếu $F \geq G$ thì $\mathbb{G}(F) \geq \mathbb{G}(G)$
Tính thuần nhất: Nếu c là hằng số thì $\mathbb{G}(cF) = c\mathbb{G}(F)$
Tính lõm: $(\mathbb{G}(F)+\mathbb{G}(G))/2 \leq \mathbb{G}((F+G)/2)$. Dấu bằng xảy ra khi và chỉ khi F và G tỷ lệ thuận với nhau, tức là tồn tại một hằng số dương c sao cho $G = cF$ hầu khắp mọi nơi.

Ghi chú 2.6. Tính lõm của giá trị kỳ vọng hình học chính là cơ sở của nguyên tắc **đa dạng hóa tài sản** (diversification) trong đầu tư: Bằng cách đa dạng hóa tài sản (đầu tư một phần vào F và một phần vào G , thay vì chỉ đầu tư vào F hay chỉ đầu tư vào G) có thể làm tăng giá trị kỳ vọng hình học của danh mục đầu tư (ít ra là trong trường hợp F và G có cùng kỳ vọng hình học về tăng trưởng).

Bài tập 2.19. Chứng minh bất đẳng thức $(\mathbb{G}(F)+\mathbb{G}(G))/2 \leq \mathbb{G}((F+G)/2)$, cho trường hợp không gian xác suất là một không gian hữu hạn phần tử có phân bố xác suất đều.

2.4 Phương sai, độ lệch chuẩn, và các moment

2.4.1 Phương sai và độ lệch chuẩn

Định nghĩa 2.16. Độ lệch chuẩn *(standard deviation) của một biến ngẫu nhiên X là*

$$\sigma(X) = \sqrt{\mathbb{E}((X - \mathbb{E}(X))^2)}. \quad (2.43)$$

Phương sai *(variance) của X, ký hiệu là $var(X)$, chính là bình phương của độ lệch chuẩn của X, tức là bằng $\mathbb{E}((X - \mathbb{E}(X))^2)$.*

Sử dụng tính tuyến tính của giá trị kỳ vọng, ta có thể biến đổi công thức của phương sai như sau: $\mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2 - 2\mathbb{E}(X).X + \mathbb{E}(X)^2) = \mathbb{E}(X^2) - 2\mathbb{E}(X).\mathbb{E}(X) + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. Như vậy, ta có công thức sau:

$$var(X) = \sigma(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \quad (2.44)$$

Độ lệch chuẩn có tính thuần nhất bậc một: $\sigma(cX) = c\sigma(X)$, còn phương sai thì thuần nhất bậc hai: $var(cX) = \sigma(cX)^2 = c^2 var(X)$. Ý nghĩa của độ lệch chuẩn là: nó là thước đo độ lệch của các giá trị của X so với giá trị trung bình của nó. Định nghĩa của phương sai cho thấy nó luôn luôn lớn hơn hoặc bằng 0, và bằng 0 khi và chỉ khi X là hằng số hầu khắp mọi nơi, tức là nó không bị lệch đi đâu cả so với giá trị trung bình của nó.

Câu hỏi cho những người tò mò: Tại sao người ta lại hay dùng phương sai và độ lệch chuẩn làm thước đo cho độ lệch giữa các giá trị của một biến ngẫu nhiên X với giá trị kỳ vọng của nó, chứ không dùng một đại lượng kiểu như $\mathbb{E}(|X - \mathbb{E}(X)|)$?

Ví dụ 2.19. Nếu F nhận hai giá trị a và $-a$ ($a > 0$), mỗi giá trị với xác suất 50%, thì giá trị kỳ vọng của F là 0, phương sai của F là $a^2.50\% + (-a)^2.50\% = a^2$, và độ lệch chuẩn chính là a.

Ví dụ 2.20. Nếu F có phân bố normal $\mathcal{N}(\mu, \sigma^2)$, thì giá trị kỳ vọng của F chính là μ, còn độ lệch chuẩn của F chính là σ. (Bài tập: chứng minh điều đó bằng các biến đổi tích phân, xuất phát từ công thức $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})dx = 1$)

Ghi chú 2.7. Đối với các biến ngẫu nhiên với vô hạn các giá trị, thì các đại lượng đặc trưng của chúng như kỳ vọng, phương sai, và các đại lượng khác, không phải lúc nào cũng tồn tại hay hữu hạn. Ví dụ, phân bố xác suất rời rạc $P(k) = C/k^2$ với mọi $k \in \mathbb{N}$, với $C = 1/(\sum 1/n^2) = 6/\pi^2$, không có kỳ vọng và sai phương hữu hạn. Ta chỉ sử dụng các đại lượng đặc trưng khi chúng tồn tại và hữu hạn.

Bài tập 2.20. Chứng minh rằng:

i) Độ lệch chuẩn của phân bố hình học với tham số p ($P(k) =$

$p(1-p)^{k-1}$ với mọi $k \in \mathbb{N}$) là $\sigma = \dfrac{\sqrt{1-q}}{q}$.
ii) Độ lệch chuẩn của phân bố Poisson với tham số λ ($P(k) = e^{-\lambda}.\lambda^k/k!$ với mọi $k \in \mathbb{Z}_+$) là $\sigma = \sqrt{\lambda}$.

Bài tập 2.21. Giả sử X là một biến ngẫu nhiên với $\mathbb{E}(X) = 2/3$, và có phân bố xác suất liên tục với hàm mật độ ρ_X có dạng sau: $\rho_X(x) = ax^2 + b$ nếu $0 < x < 1$, và $\rho_X(x) = 0$ ở những điểm còn lại. Hãy tính a, b, và $var(X)$.

Bài tập 2.22. Một phòng thí nghiệm phải kiểm tra một lượng N rất lớn các mẫu máu người (mỗi mẫu của 1 người) để tìm ra các mẫu có chứa một loại kháng thể X. Thay vì xét nghiệm từng mẫu một, người ta làm như sau: Chia các mẫu thành từng nhóm, mỗi nhóm có k mẫu. Trộn các mẫu máu trong cùng một nhóm với nhau (lấy một ít máu từ mỗi mẫu) để được 1 mẫu hỗn hợp, rồi xét nghiệp mẫu hỗn hợp đó. Nếu kết quả xét nghiệm là âm tính (mẫu hỗn hợp không có kháng thể X) thì coi như cả k mẫu trong nhóm đều không có kháng thể X, còn nếu mẫu hỗn hợp có kháng thể X, thì làm tiếp k xét nghiệm, mỗi xét nghiệm cho từng mẫu của nhóm. Giả sử xác suất để 1 mẫu máu có kháng thể X là một số p, và các mẫu máu độc lập với nhau. Gọi S là tổng số lần phải xét nghiệm.
i) Xác suất để một mẫu máu hỗn hợp có chứa kháng thể X là bao nhiêu?
ii) Tính kỳ vọng và phương sai của S, khi tổng số mẫu máu phải kiểm tra là $N = km$.
iii) Với những giá trị nào của p thì tồn tại một số k thích hợp nào đó sao cho phương pháp xét nghiệm trên tiết kiệm được số lần xét nghiệm (kỳ vọng của S nhỏ hơn N)? Tìm giá trị của k tối ưu, như là

hàm của p.

2.4.2 Các moment của một biến ngẫu nhiên

Định nghĩa 2.17. *Nếu X là một biến ngẫu nhiên, và k là một số tự nhiên, thì đại lượng $\mathbb{E}(X^k)$ được gọi là* **moment** *(hay* **mô men***) bậc k của X, và đại lượng $\mathbb{E}((X - \mathbb{E}(X))^k)$ được gọi là* **moment trung tâm** *bậc k của X.*

Ghi chú 2.8. Có nhiều từ thuật ngữ gốc nước ngoài, mà trong tiếng Việt không có từ "thuần Việt" tương ứng, chỉ dịch phiên âm, ví dụ như mô men (moment), véc tơ (vector), mô đun (module), v.v. Trong những trường hợp như vậy, ở đây chúng ta sẽ để nguyên từ theo tiếng Anh, thay vì dùng phiên âm tiếng Việt.

Như phía trên chúng ta đã thấy, moment bậc 1 của X chính là giá trị kỳ vọng của nó, moment trung tâm bậc 1 của X thì luôn bằng 0, moment trung tâm bậc 2 của X chính là phương sai của nó, và nó có thể được biểu diễn qua các moment của X theo công thức:

$$\mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \qquad (2.45)$$

Tương tự như vậy, các moment trung tâm bậc cao hơn của X cũng có thể khai triển dưới dạng đa thức của các moment của X.

Nếu ký hiệu P_X là phân bố xác suất trên $\mathbb{R}$ của X, thì ta có thể viết moment bậc k của X theo công thức sau:

$$\mathbb{E}(X^k) = \int_{x \in \mathbb{R}} x^k dP_X. \qquad (2.46)$$

Nếu như phân bố xác suất P_X la một phân bố xác suất liên tục với hàm mật độ ρ_X thì ta có thể viết:

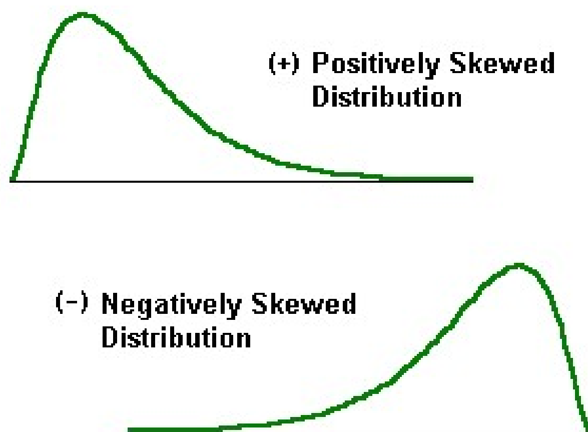
$$\mathbb{E}(X^k) = \int_{-\infty}^{+\infty} x^k \rho_X(x) dx. \tag{2.47}$$

Các moment của một biến ngẫu nhiên cho ta các thông tin về dáng điệu của phân bố xác suất của biến ngẫu nhiên đó. Ví dụ, nếu moment trung tâm bậc 2 nhỏ, thì có nghĩa là các giá trị của X nói chung ít bị sai lệch so với giá trị kỳ vọng của nó, hay nói cách khác phần lớn xác suất của phân bổ xác suất của X tập trung trong một khoảng nhỏ xung quanh điểm giá trị kỳ vọng. Ngược lại, nếu moment trung tâm bậc 2 lớn, thì phân bố xác suất của X nói chung sẽ "dàn trải" hơn ra xa điểm giá trị kỳ vọng.

Moment trung tâm bậc 3 của X được gọi là **hệ số bất đối xứng** (skewness), hay còn có thể gọi là **độ xiên** của phân bố xác suất của X: Nếu X có phân bố xác suất đối xứng quanh điểm giá trị kỳ vọng (có nghĩa là X và $2\mathbb{E}(X) - X$ có cùng phân bố xác suất), thì moment trung tâm bậc 3 của nó bằng 0. Nếu như moment trung tâm bậc 3 lớn hơn 0 thì phân bố xác suất của X được gọi là *xiên về bên phải*, còn nếu moment trung tâm bậc 3 nhỏ hơn 0 thì phân bố xác suất của X được gọi là *xiên về bên trái*.

Ví dụ 2.21. Moment trung tâm bậc 3 của một phân bố normal bằng 0.

Ví dụ 2.22. Giả sử có một biến ngẫu nhiên X với phân bố xác suất rời rạc sau: $P(X = -2) = 1/2, P(X = 1) = 1/4, P(X = 3) = 1/4$. Khi đó giá trị kỳ vọng của X bằng 0, moment trung tâm bậc 3 của X bằng moment bậc 3 của X và bằng: $(1/2).(-2)^3 + (1/4).1^3 + (1/4).3^3 =$



Hình 2.8: Phân bố bất đối xứng

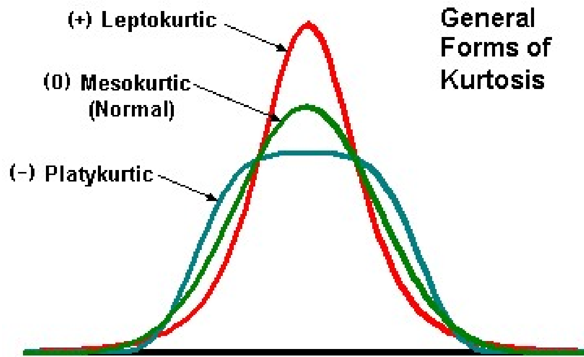
$3 > 0$. Đồ thị phân bố xác suất của X (với 3 đoạn thẳng nhô lên ở 3 điểm -2,1,3 trên trục hoành) bị "lệch về bên phải" so nếu lấy điểm giá trị kỳ vọng $(= 0)$ làm trung điểm.

Moment trung tâm bậc 4 của X liên quan đến cái gọi là kurtosis[(3)] của X. Theo định nghĩa, **kurtosis** (hay còn gọi là **hệ số nhọn**) của một biến ngẫu nhiên là đại lượng

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3, \tag{2.48}$$

trong đó μ_4 là moment trung tâm bậc 4, còn σ là độ lệch chuẩn. Tỷ lệ μ_4/σ^4 được gọi là **moment chuẩn hóa** bậc 4. Lý do của việc chuẩn hóa này là: các moment chuẩn hóa của các phân bố normal đều là hằng số và không phụ thuộc vào độ lệch chuẩn. Moment chuẩn hóa

[(3)]kurtosis là một từ gốc tiếng Hy lạp, chỉ độ nhọn

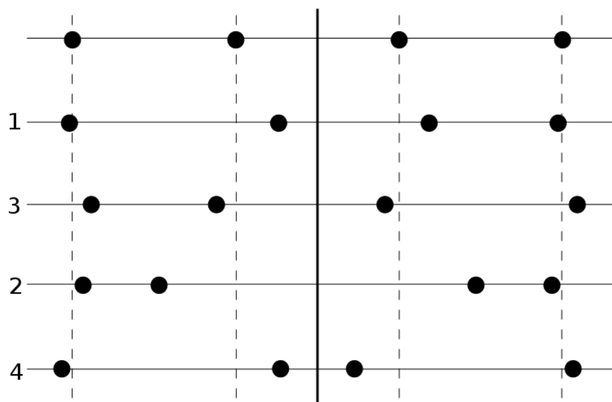


Hình 2.9: Kurtosis

của bậc 4 của một phân bố normal chính bằng 3, bởi vậy kurtosis của một phân bố normal bằng 0. Khi một phân bố xác suất có kurtosis dương (phân bố như vậy gọi là phân bố **leptokurtic** hay **nhọn vượt chuẩn**) thì có nghĩa là nó "nhọn" hơn phân bố normal có cùng độ lệch chuẩn, còn khi kurtosis âm (phân bố như vậy gọi là phân bố **platykurtic**) thì có nghĩa là nó "bẹt" hơn phân bố normal có cùng độ lệch chuẩn. Nếu kurtosis bằng 0 thì phân bố được gọi là **mesokurtic**. (Xem hình 2.9).

Ví dụ 2.23. Hình 2.10 là ví dụ minh họa về việc dịch chuyển 4 điểm a, b, c, d của một phân bố xác suất đều rời rạc $P(a) = P(b) = P(c) = P(d) = 1/4$, từ vị trí ban đầu $a = -3, b = -1, c = 1, d = 3$, sao cho làm tăng 1 trong 4 moment bậc 1, bậc 2, bậc 3, bậc 4 trong khi giữ nguyên 3 moment còn lại.

Tất nhiên, nếu hai biến ngẫu nhiên có cùng phân bố xác suất trên $\mathbb{R}$, thì tất cả các moment của chúng đều bằng nhau. Điều ngược



Hình 2.10: Thay đổi các moment bậc 1 đến bậc 4

có đúng không, hay nói cách khác, dãy các moment $\mathbb{E}(X^k), k = 1, 2, 3, \ldots$ của một biến ngẫu nhiên xác định hoàn toàn phân bố xác suất của biến ngẫu nhiên đó không? Đây là một câu hỏi toán học thú vị. Có những ví dụ về các phân bố xác suất liên tục khác nhau nhưng có tất cả các moment như nhau. Tuy nhiên, trong trường hợp các không gian xác suất chỉ có hữu hạn phần tử (mà thực ra tất cả các vấn đề trong thực tế đều chỉ có hữu hạn các khả năng xảy ra, và các mô hình liên tục với vô hạn khả năng chỉ là các mô hình mô phỏng gần đúng), thì ta có:

Mệnh đề 2.14. *Nếu X và Y là hai biến ngẫu nhiên chỉ nhận một số hữu hạn các giá trị, và có $\mathbb{E}(X^k) = \mathbb{E}(Y^k)$ với mọi $k \in \mathbb{N}$, thì phân bố xác suất của chúng trên $\mathbb{R}$ bằng nhau.*

Bài tập 2.23. Chứng minh mệnh đề trên.

Bài tập 2.24. Tính kỳ vọng và các moment của phân bố mũ với tham

số λ.

2.4.3 Bất đẳng thức Chebyschev và bất đẳng thức Markov

Những bất đẳng thức tương đối đơn giản sau đây của Chebyschev và Markov liên quan đến các moment sẽ có ích trong việc đánh giá phân bố xác suất của các biến ngẫu nhiên.

Định lý 2.15. *(Bất đẳng thức Chebyschev cho kỳ vọng) Với mọi biến ngẫu nhiên X chỉ nhận các giá trị không âm, và mọi số dương $a > 0$ ta có*

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}. \tag{2.49}$$

Chứng minh. Gọi X_a là biến ngẫu nhiên sau: $X_a = a$ khi $X \geq a$ và $X_a = 0$ khi $X < a$. Khi đó $X \geq X_a$, và X_a chỉ nhận hai giá trị 0 và a. Bởi vậy

$$\mathbb{E}(X) \geq \mathbb{E}(X_a) = 0.P(X_a = 0) + a.P(X_a = a) = a.P(X \geq a),$$

từ đó suy ra điều phải chứng minh. □

Định lý 2.16. *(Bất đẳng thức Markov cho các moment tuyệt đối) Với mọi biến ngẫu nhiên X, số dương $a > 0$, và số tự nhiên k, ta có*

$$P(|X| \geq a) \leq \frac{\mathbb{E}(|X|^k)}{a^k}. \tag{2.50}$$

Chứng minh. Suy ra từ bất đẳng thức Chebyschev cho biến ngẫu nhiên $|X|^k$ và hằng số a^k. □



Hình 2.11: Pafnouti Lvovitch Chebyschev (1821-1894)

Định lý 2.17. *(Bất đẳng thức Chebyschev cho phương sai) Nếu X là một biến ngẫu nhiên có phương sai $var(X)$ hữu hạn và $a > 0$ bất kỳ, ta có*

$$P(|X - \mathbb{E}(X)| \geq a) \leq \frac{var(X)}{a^2}. \quad (2.51)$$

Chứng minh. Suy ra từ bất đẳng thức Markov cho biến ngẫu nhiên $X - \mathbb{E}(X)$ và cho $k = 2$. □

Ghi chú 2.9. Pafnouti Lvovitch Chebyschev (1821-1894) là một nhà



Hình 2.12: Andrei Andreevitch Markov (1856-1922)

toán học người Nga. Ngoài lý thuyết xác suất, ông ta còn nghiên cứu nhiều về số học và đại số. Các đa thức U_n bậc n thỏa mãn $U_n(\cos(x)) = \frac{\sin((n+1)x)}{\sin x}$ được gọi là *đa thức Chebyschev*, và chúng xuất hiện nhiều trong toán học và ứng dụng. Andrei Andreevitch Markov (1856-1922) cũng là một nhà toán học người Nga, và là học trò của Chebyschev. Các *xích Markov* (Markov chains) đặc biệt quan

trọng trong lý thuyết xác suất về các quá trình ngẫu nhiên (stochastic processes). Các quá trình ngẫu nhiên nằm ngoài khuôn khổ của cuốn sách này, nhưng sẽ được bàn đến trong một cuốn sách tiếp theo.

2.5 Hàm đặc trưng, hàm sinh, và biến đổi Laplace

Thay vì xét các moment $\mathbb{E}(X^k)$ của một biến ngẫu nhiên X, ta có thể xét các giá trị đặc trưng dạng $\mathbb{E}(\exp(yX))$ trong đó y là một tham số nào đó. Khi ta biến đổi y trong một miền nào đó trên $\mathbb{R}$ hoặc $\mathbb{C}$, sẽ ta được một hàm các giá trị đặc trưng của X. Sự liên quan giữa hàm này và các moment được thể hiện qua đẳng thức sau (xảy ra nếu như ta có các điều kiện về hội tụ):

$$M_X(y) = \mathbb{E}(\exp(yX)) = \mathbb{E}(\sum_k (y^k/k!).X^k) = \sum_k \mathbb{E}(X^k).(y^k/k!) \tag{2.52}$$

Hàm $M_X(y) = \mathbb{E}(\exp(yX))$ được gọi là **hàm sinh moment** của X.

2.5.1 Hàm đặc trưng

Trong biểu thức $M_X(y) = \mathbb{E}(\exp(yX))$, nếu ta lấy $y = is$, (ở đây $i = \sqrt{-1}$) ,với $s \in \mathbb{R}$, thì ta có $\exp(yX) = \exp(isX) = \cos(sX) + i\sin(sX)$ là một biến ngẫu nhiên bị chặn (có giá trị tuyệt đối bằng 1), và ta có thể yên tâm về sự tồn tại của $\mathbb{E}(\exp(isX))$. Từ đó có định nghĩa sau:

Định nghĩa 2.18. Hàm đặc trưng *của một biến ngẫu nhiên thực X là hàm $\Phi_X : \mathbb{R} \to \mathbb{C}$ được cho bởi công thức*

$$\Phi_X(s) = \mathbb{E}(\exp(isX)) = \int_{x\in\mathbb{R}} e^{isx} dP_X. \tag{2.53}$$

Ví dụ 2.24. Hàm đặc trưng của một số phân bố xác suất quen thuộc:
i) Hàm đặc trưng của một hằng số c (tức là biến ngẫu nhiên chỉ nhận mỗi giá trị c) là $\Phi_c(s) = e^{ics}$.
ii) Hàm đặc trưng của phân bố nhị thức với các tham số n, p là hàm $(1 - p + pe^{is})^n$.
iii) Hàm đặc trưng của phân bố xác suất đều trên một đoạn thẳng $[a, b]$ là hàm $\dfrac{e^{ibs} - e^{-ias}}{i(b-a)s}$.
iv) Hàm đặc trưng của phân bố xác suất mũ với tham số 1 (với mật độ $\rho(x) = e^{-x}$ khi $x > 0$) là hàm $\dfrac{1}{1 - is}$.
v) Hàm đặc trưng của phân bố xác suất normal chuẩn tắc $\mathcal{N}(0, 1)$ là hàm $\Phi(s) = \exp(-s^2/2)$.
(Bài tập: Hãy suy ra các công thức trên từ định nghĩa của hàm đặc trưng và của các phân bố xác suất).

Định lý 2.18. *Một số tính chất của hàm đặc trưng:*
i) $\Phi_X(0) = 1$
ii) $|\Phi_X(s)| \leq 1$ *với mọi* $s \in \mathbb{R}$
iii) Nếu $Y = aX + b$ *với* a, b *là các hằng số, thì* $\Phi_Y(s) = e^{\sqrt{-1}bs}\Phi_X(as)$.
iv) Φ_X *liên tục đều trên* $\mathbb{R}$.
v) Nếu $\mathbb{E}(|X|^k) < \infty$ *với một số tự nhiên* k *nào đó, thì hàm đặc trưng* Φ_X *khả vi liên tục* k *lần trên* $\mathbb{R}$*, và*

$$\mathbb{E}(X^k) = \frac{1}{(\sqrt{-1})^k}.\Phi_X^{(k)}(0), \tag{2.54}$$

trong đó $\Phi_X^{(k)}$ *là ký hiệu đạo hàm bậc* k *của* Φ_X.

Chứng minh. Ba tính chất đầu tiên tương đối hiển nhiên, suy ra ngay từ định nghĩa. Tính chất thứ tư là bài tập dành cho những bạn đọc quen với khái niệm liên tục đều. Để chứng minh tính chất cuối cùng, chúng ta nhớ rằng phép lấy giá trị kỳ vọng là một phép lấy giá trị trung bình, có thể hiểu như là một phép lấy tổng (của một chuỗi), và do đó nó giao hoán với phép lấy đạp hàm (khi một số điều kiện hội tụ nào đó được thỏa mãn). Áp dụng nguyên tắc giao hoán đó vào định nghĩa của hàm đặc trưng, ta có đạo hàm bậc k của hàm đặc trưng là:

$$\Phi_X^{(k)}(s) = \frac{d^k}{ds^k}\Phi_X(s) = \mathbb{E}(\frac{d^k}{ds^k}\exp(isX)) \\ = \mathbb{E}((iX)^k\exp(isX)) = i^k\mathbb{E}(X^k\exp(isX)). \quad (2.55)$$

Đặt $s = 0$, ta được $\Phi_X^{(k)}(0) = i^k\mathbb{E}(X^k)$. □

2.5.2 Tìm lại phân bố xác suất từ hàm đặc trưng

Chúng ta có công thức giới hạn sau đây, cho phép tìm lại được phân bố xác suất từ hàm đặc trưng của nó:

Định lý 2.19. *Gọi* P_X *là phân bố xác suất của một biến ngẫu nhiên* X *tùy ý, và* Φ_X *là hàm đặc trưng của nó. Khi đó với mọi* $a, b \in \mathbb{R}$, $a < b$, *ta có:*

$$\lim_{R\to\infty}\frac{1}{2\pi}\int_{-R}^{R}\frac{e^{-ias}-e^{-ibs}}{is}\Phi_X(s)ds = P_X(]a,b[) + \frac{P_X(a)+P_X(b)}{2}. \quad (2.56)$$

Chúng ta sẽ chấp nhận định lý trên mà không chứng minh. Nếu bạn đọc đã biết qua về giải tích Fourier thì có thể tự chứng minh nó không quá khó khăn (nó tương tự như định lý Dirichlet cho chuỗi Fourier). Nếu không thì có thể xem chẳng hạn trong Chương 9 của quyển sách của Koralov và Sinai [5].

Trong trường hợp X có phân bố xác suất liên tục với hàm mật độ ρ_X, thì ta có thể viết

$$\Phi_X(s) = \int_{-\infty}^{+\infty} e^{isx} \rho_X(x) dx. \tag{2.57}$$

Trong giải tích, phép tính trên gọi là phép biến đổi Fourier. Có nghĩa là, hàm đặc trưng chính là **biến đổi Fourier** của hàm mật độ.

Chia cả hai vế của công thức (2.56) cho $b - a$, và cho b tiến tới a, ta được công thức sau, gọi là phép **biến đổi ngược Fourier**, để tính hàm mật độ từ hàm đặc trưng:

$$\rho_F(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx} \Phi_F(s) ds. \tag{2.58}$$

Trong trường hợp X là biến ngẫu nhiên nguyên (chỉ nhận giá trị trong $\mathbb{Z}$), thì hàm đặc trưng Φ_X của X chính là **chuỗi Fourier** với các hệ số là các xác suất $P_X(k) = P(X = k)$, $k \in \mathbb{Z}$:

$$\Phi_X(s) = \sum_{k \in \mathbb{Z}} P_X(k) \exp(iks), \tag{2.59}$$

và ta có thể tính ra $P_X(k)$ từ Φ_X theo công thức quen thuộc để tính các hệ số của một chuỗi Fourier:

$$P_X(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-iks} \Phi_X(s) ds \tag{2.60}$$

Ghi chú 2.10. Joseph Fourier (1768–1830) là nhà toán học và vật lý Pháp. Trong khoảng thời gian 1798–1801 Fourier đi theo Napoléon, cùng với 35000 lính Pháp và một đoàn các nhà khoa học, sang chinh chiến ở Ai Cập (Egypt) và tìm hiểu nền văn minh Ai Cập. Khi ở Ai Cập, Fourier trở thành người điều hành Viện Hàn lâm Ai Cập do Napoléon lập ra, và sau đó điều hành luôn cả các công việc hành chính và ngoại giao ở Ai Cập, gần như là quan toàn quyền. Fourier tỏ ra rất có tài về chính trị và ngoại giao, có thể đàm phán, hòa giải các bên đối lập. Sau khi Pháp đầu hàng Anh ở Ai Cập năm 1801 và Fourier trở về Pháp, được cử làm tỉnh trưởng (préfet) vùng Isère. Trong thời gian ở Ai Cập, Fourier phát minh ra chuỗi Fourier, khi nhìn thấy các lớp sóng cát (dunes) ở sa mạc. Chuỗi Fourier và biến đổi Fourier là một thứ *công cụ vạn năng*, không chỉ quan trọng trong xác suất, mà còn xuất hiện khắp nơi trong toán học và vật lý.

Trong trường hợp tổng quát, một phân bố xác suất cũng được xác định một cách duy nhất bởi hàm đặc trưng của nó:

Định lý 2.20. *Hai biến ngẫu nhiên có cùng phân bố xác suất khi và chỉ khi chúng có cùng hàm đặc trưng.*

Chứng minh. Giả sử hai phân bố xác suất P_X và P_Y có cùng hàm đặc trưng Φ. Công thức (2.56) dẫn đến:

$$P_X(]a,b[) + \frac{P_X(a) + P_X(b)}{2} = P_Y(]a,b[) + \frac{P_Y(a) + P_Y(b)}{2}$$

với mọi $a < b$. Ta có thể chọn a và b là những điểm liên tục của $\mathcal{F}_X$ và $\mathcal{F}_Y$, rồi cho a tiến tới $-\infty$, ta được: $\mathcal{F}_X(b) = \mathcal{F}_Y(b)$ tại mọi điểm b mà là điểm liên tục của cả $\mathcal{F}_X$ và $\mathcal{F}_Y$. Giả sử $x \in \mathbb{R}$ là một điểm tùy ý.



Hình 2.13: Joseph Fourier (1768–1830)

Nhắc lại rằng số điểm gián đoạn của một hàm phân phối xác suất trên $\mathbb{R}$ là không quá đếm được. Vì thế tồn tại một dãy các điểm $x_n > x$ sao cho x_n tiến tới x khi n tiến tới vô cùng, và x_n là điểm liên tục của $\mathcal{F}_X$ và $\mathcal{F}_Y$ với mọi n. Nhắc lại rằng, các hàm phân phối xác suất có tính chất liên tục bên phải. Do đó ta có: $\mathcal{F}_X(x) = \lim_{n\to\infty} \mathcal{F}_X(x_n) = \lim_{n\to\infty} \mathcal{F}_Y(x_n) = \mathcal{F}_Y(x)$. Như vậy, hai hàm phân phối xác suất $\mathcal{F}_X$ và $\mathcal{F}_Y$ trùng nhau, do đó hai phân bố xác suất P_X và P_Y cũng trùng nhau. □

Bài tập 2.25. Ta sẽ gọi một biến ngẫu nhiên X là *đối xứng* nếu như

X và $-X$ có cùng phân bố xác suất. Hãy xây dựng những ví dụ biến ngẫu nhiên đối xứng, và chứng minh rằng một biến ngẫu nhiên là đối xứng khi và chỉ khi hàm đặc trưng Φ_X của nó là một hàm thực (tức là $\Phi_X(s) \in \mathbb{R}$ với mọi $s \in \mathbb{R}$).

2.5.3 Hàm sinh xác suất và biến đổi Laplace

Trong biểu thức $\mathbb{E}(\exp(yX))$, nếu đặt $y = \ln z$, thì ta được hàm sau, gọi là **hàm sinh xác suất**:

$$G_X(z) = \mathbb{E}(z^X) \tag{2.61}$$

Hàm sinh xác suất hay được dùng khi mà các giá trị của biến ngẫu nhiên đều là số nguyên không âm. Khi đó hàm sinh xác suất có dạng đa thức hoặc chuỗi Taylor có bán kính hội tụ lớn hơn hoặc bằng 1:

$$G_X(z) = \sum_k P_X(k).z^k, \tag{2.62}$$

và ta có $P(X = k) = \frac{1}{k!} \frac{d^k G_X(z)}{dz^k}\Big|_{z=0}$ với mọi $k \in \mathbb{Z}_+$.

Từ quan điểm của giải tích phức, hàm đặc trưng $\Phi_X(s)$ và hàm sinh $G_X(z)$ gần như là một, có thể chuyển từ hàm này sang hàm kia bằng cách đổi biến. Bởi vậy, tất nhiên các moment của một biến ngẫu nhiên cũng có thể suy ra được từ hàm sinh xác suất của biến ngẫu nhiên đó. Ta có định lý sau:

Định lý 2.21. *Giả sử X là một biến ngẫu nhiên với hàm sinh xác suất G . Khi đó:*

1) $\mathbb{E}(X) = G'(1)$

2) $var(X) = \sigma^2(X) = G''(1) + G'(1) - (G'(1))^2$
3) $\mathbb{E}(X(X-1)\dots(X-k+1)) = G^{(k)}(1)$ *với mọi* $k \in \mathbb{N}$. *Ở đây* $G^{(k)}$ *là đạo hàm bậc* k *của* G.

Ví dụ 2.25. Hàm sinh xác suất của một biến ngẫu nhiên X với phân bố Poisson với tham số λ là hàm $G_X(z) = \exp((z-1)\lambda)$ Thật vậy, ta có: $G_X(z) = \mathbb{E}(z^X) = \sum_k z^k P(X = k) = \sum_k e^{-\lambda}\lambda^k z^k/k! = e^{-\lambda}e^{\lambda z} = e^{\lambda(z-1)}$. Từ đó suy ra $\mathbb{E}(X) = G'_X(1) = \lambda, G''_X(1) = \lambda^2$ và $var(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Trong trường hợp biến ngẫu nhiên X chỉ nhận các giá trị thực không âm, người ta hay dùng **hàm Laplace** $L_X(t) : \mathbb{R}_+ \to \mathbb{R}$, nhận được từ biểu thức $\mathbb{E}(\exp(yX))$ bằng cách đặt $t = -y$:

$$L_X(t) = \mathbb{E}(\exp(-tX)). \tag{2.63}$$

Ở đây ta coi biến t nằm trong tập các số thực không âm. Với giả sử rằng F chỉ nhận các giá trị không âm, ta luôn có $0 < (\exp(-tF)) \leq 1$, từ đó suy ra các giá trị của $L_F(t)$ là số dương và bị chặn trên bởi 1.

Trong trường hợp F có phân bố xác suất liên tục với hàm mật độ ρ_F thỏa mãn điều kiện $\rho_F(x) = 0$ với mọi $x < 0$ (có nghĩa là F không nhận các giá trị âm), thì ta có

$$L_F(t) = \int_0^\infty \exp^{-tx} \rho_F(x)dx, \tag{2.64}$$

và hàm $L_F(t)$ được gọi là **biến đổi Laplace** của hàm mật độ $\rho_F(x)$.

Tương tự như đối với hàm sinh và hàm đặc trưng, các đạo hàm của hàm $L_F(t)$ tại điểm $t = 0$ cũng cho ta các moment của F.



Hình 2.14: Pierre-Simon Laplace (1949-1827)

Ghi chú 2.11. [(4)] Pierre-Simon Laplace (1749-1827) là nhà toán học, thiên văn học và vật lý người Pháp, một trong những nhà khoa học có thế lực nhất ở châu Âu thời đại ông ta. Ông ta nghiên cứu rất nhiều thứ, từ xác suất (định lý giới hạn trung tâm, biến đổi Laplace) đến giải tích điều hòa, cơ học, âm thanh, truyền nhiệt, các thiên thể, v.v. Laplace chính là người đặt ra giả thuyết về lỗ đen (black hole) và về sự co lại do trọng lượng (gravitational collapse) trong vật lý thiên

[(4)]Xem wikipedia: http://fr.wikipedia.org/wiki/Pierre-Simon_Laplace.

văn. Laplace còn có tham vọng về chính trị, là thành viên của thượng nghị viện. Có lúc làm Bộ trưởng Bộ nội vụ dưới thời Napoléon, nhưng sau 6 tuần thì bị cách chức vì không được việc. Laplace bị nhiều người cùng thời không ưa vì tính bạc bẽo, ích kỷ, có khi còn vơ cả công trình của người khác thành của mình, và thay đổi quan điểm chính trị như chong chóng "theo chiều gió". Nhưng về mặt khoa học, Laplace là một con người vĩ đại của thế kỷ 18-19. Biến đổi Laplace được gọi như vậy là do Laplace đưa vào để nghiên cứu xác suất, cùng với hàm sinh xác suất. Biến đổi Laplace còn xuất hiện ở nhiều nơi khác trong vật lý và toán học. Leonhard Euler (1707–1783) có lẽ là người đầu tiên nghĩ ra biến đổi này.

Bài tập 2.26. Chứng minh rằng hàm sinh xác suất của một biến ngẫu nhiên với phân bố hình học với tham số p là hàm $G(z) = \frac{pz}{1-z+pz}$. Từ đó suy ra kỳ vọng và phương sai của phân bố hình học.

Bài tập 2.27. Tính hàm sinh xác suất và hàm Laplace của phân bố nhị thức với các tham số n, p.

Bài tập 2.28. Chứng minh định lý 2.21 cho trường hợp F nhỉ nhận một số hữu hạn các giá trị.

Chương 3

Vector ngẫu nhiên

3.1 Vector ngẫu nhiên

3.1.1 Phân bố xác suất đồng thời

Nếu ta có hai biến ngẫu nhiên $X, Y : (\Omega, P) \to \mathbb{R}$, thì ta có thể xét chúng cùng một lúc với nhau như là một biến ngẫu nhiên với giá trị trong $\mathbb{R}^2$:

$$\mathbf{X} = (X, Y) : (\Omega, P) \to \mathbb{R}^2. \tag{3.1}$$

Một biến ngẫu nhiên $\mathbf{X}$ với giá trị trong $\mathbb{R}^2$ còn được gọi là một **vector ngẫu nhiên**(1) 2 chiều. Tương tự như vậy, nếu ta có n biến ngẫu nhiên với giá trị thực, ta có thể xét chúng cùng một lúc với nhau như là một biến ngẫu nhiên với giá trị trong $\mathbb{R}^n$, và gọi nó là một vector ngẫu nhiên n chiều.

(1) Tiếng Việt phiên âm chữ vector thành **véc tơ**, nhưng ở đây chúng ta sẽ để nguyên chữ vector cho tiện.

Định nghĩa 3.1. *Một vector ngẫu nhiên n chiều $\mathbf{X} = (X_1, \ldots, X_n) : (\Omega, P) \to \mathbb{R}^n$ xác định trên $\mathbb{R}^n$ một phân bố xác suất cảm sinh qua push-forward từ phân bố xác suất trên Ω. Phân bố xác suất trên $\mathbb{R}^n$ này được gọi là* **phân bố xác suất** *của $\mathbf{X}$, hay còn được gọi là* **phân bố xác suất đồng thời** *của các biến ngẫu nhiên $X_1, \ldots, X_n$. Hàm $\mathcal{F}_{\mathbf{X}} : \mathbb{R}^n \to [0, 1]$ cho bởi công thức*

$$\mathcal{F}_{\mathbf{X}}(x_1, \ldots, x_n) := P(X_1 \leq x_1, \ldots, X_n \leq x_n) \tag{3.2}$$

được gọi là **hàm phân phối xác suất** *của vector ngẫu nhiên $\mathbf{X}$, hay còn gọi là* **hàm phân phối xác suất đồng thời** *của các biến ngẫu nhiên $X_1, \ldots, X_n$.*

Ghi chú 3.1. Nói một cách chặt chẽ toán học, tương tự như trong trường hợp 1 chiều, trong định nghĩa vector ngẫu nhiên có điều kiện **đo được**, tức là tồn tại xác suất $P(\mathbf{X} \in U)$ với mọi tập con mở U của $\mathbb{R}^n$. Khi nói đến một phân bố xác suất trên $\mathbb{R}^n$ ta sẽ luôn coi rằng sigma-đại số của nó chính là **sigma-đại số Borel** sinh bởi các tập con mở của $\mathbb{R}^n$.

Ghi chú 3.2. Nếu hai biến ngẫu nhiên $X_1 : (\Omega_1, P_1) \to \mathbb{R}$ và $X_1 : (\Omega_2, P_1) \to \mathbb{R}$ có hai mô hình không gian xác suất khác nhau, thì trước khi có thể xét cặp (X_1, X_R) như là một vector ngẫu nhiên, ta phải thay đổi mô hình không gian xác suất, để biến X_1 và X_2 thành các biến ngẫu nhiên trên cùng một không gian xác suất. Nói cách khác, ta phải xây dựng được một không gian xác suất (Ω, P), cùng với các toàn ánh bảo toàn xác suất $\phi_1 : (\Omega, P) \to (\Omega_1, P_1)$ và $\phi_2 : (\Omega, P) \to (\Omega_2, P_2)$, sao cho thích hợp với vấn đề đang được nghiên cứu. Sau đó, thay vì xét X_1 và X_2 riêng lẻ, ta có thể xét

$\tilde{X}_1 = X_1 \circ \phi_1$ và $\tilde{X}_2 = X_2 \circ \phi_2$ cùng nhau trên Ω. Chú ý rằng, về mặt bản chất, X_1 và $\tilde{X}_1$ chẳng qua là một (và tất nhiên có cùng phân bố xác suất trên $\mathbb{R}$), nhưng được đặt trên các mô hình không gian xác suất khác nhau.

Tương tự như trong trường hợp 1 chiều, một phân bố xác suất nhiều chiều $P_{\mathbf{X}}$ (của một vector ngẫu nhiên $\mathbf{X} = (X_1, \ldots, X_n)$) được xác định duy nhất bởi hàm phân phối xác suất của nó. Ví dụ, khi $n = 2$, ta có thể tính xác suất của một hình chữ nhật nửa mở $]a, b] \times]c, d]$ trong $\mathbb{R}^2$ khi biết hàm phân phối xác suất $\mathcal{F}_{\mathbf{X}}$ qua công thức sau:

$$P_{\mathbf{X}}(]a, b] \times]c, d]) = \mathcal{F}_{\mathbf{X}}(b, d) - \mathcal{F}_{\mathbf{X}}(b, c) + \mathcal{F}_{\mathbf{X}}(a, c) - \mathcal{F}_{\mathbf{X}}(a, d), \quad (3.3)$$

còn xác suất của các miền hình chữ nhật đóng thì có thể tính qua giới hạn

$$P_{\mathbf{X}}([a, b] \times [c, d]) = \lim_{a' \to a-, c' \to c-} P_{\mathbf{X}}(]a', b] \times]c', d]). \quad (3.4)$$

Bài tập 3.1. Viết công thức tính xác suất $P_{\mathbf{F}}(]a, b] \times]c, d] \times]e, f])$ của một hình khối chữ nhật nửa mở thông qua hàm phân phối xác suất $\mathcal{F}_{\mathbf{F}}$ của một vector ngẫu nhiên 3 chiều $\mathbf{F} = (F_1, F_2, F_3)$.

Bài tập 3.2. Hai người hẹn gặp nhau vào một buổi trưa tại một điểm X. Mỗi người đi đến điểm X trong khoảng thời gian từ 12h đến 13h một cách ngẫu nhiên với phân bố đều, và nếu khi đến không thấy người kia đâu thì đợi thêm 15 phút mà vẫn không thấy thì bỏ đi. Tính xác suất để hai người gặp được nhau ở điểm X theo hẹn.

3.1.2 Các phân bố xác suất biên

Khi ta có một vector ngẫu nhiên $\mathbf{X} = (X_1, \ldots, X_n)$ với phân hàm phân phối xác suất đồng thời $\mathcal{F}_{\mathbf{X}}$, thì ta có thể tìm lại được các hàm phân phối của các phân bố xác suất $\mathcal{F}_{X_i}$ của các biến X_i qua công thức giới hạn sau:

$$\mathcal{F}_{X_i}(x) = \lim_{x_k \to \infty\ \forall k \neq i} \mathcal{F}_{\mathbf{X}}(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n). \tag{3.5}$$

Do đó, các hàm phân phối xác suất $\mathcal{F}_{X_i}$ còn được gọi là các **hàm phân phối xác suất biên** (hay còn gọi là phân phối xác suất **biên duyên**) của hàm phân phối xác suất đồng thời $\mathcal{F}_{\mathbf{X}}$, và các phân bố xác suất P_{X_i} được gọi là các **phân bố xác suất biên** (marginal distributions) của phân bố xác suất đồng thời $P_{\mathbf{X}}$.

Mệnh đề 3.1. *Các phép chiếu tự nhiên* $\phi_i : (\mathbb{R}^n, P_{\mathbf{X}}) \to (\mathbb{R}, P_{X_i})$*,* $\phi_i(x_1, \ldots, x_n) = x_i$*, là các ánh xạ bảo toàn xác suất.*

Chứng minh của mệnh đề trên suy ra trực tiếp từ các định nghĩa.

Ví dụ 3.1. Nếu X là một biến ngẫu nhiên rời rạc với các giá trị x_i và Y là một biến ngẫu nhiên rời rạc với các giá trị y_j, thì việc phép chiếu thứ nhất bảo toàn xác suất có nghĩa là

$$P(X = x_i) = P(\cup_j (X = x_i, Y = y_j)) = \sum_j P(X = x_i, Y = y_j). \tag{3.6}$$

Chú ý rằng các phân bố xác suất biên được xác định duy nhất bởi phân bố xác suất đồng thời, nhưng điều ngược lại nói chung không đúng: Nếu ta biết phân bố xác suất của hai biến ngẫu nhiên X, Y thì không có nghĩa là ta biết phân bố xác suất đồng thời của chúng.

Ví dụ 3.2. Giả sử ta biết rằng P_X là phân bố Bernoulli với $P_X(0) = 1 - p, P_X(1) = p$, và P_Y cũng là phân bố Bernoulli với $P_Y(0) = 1-q, P_Y(1) = q$. Khi đó ta biết rằng $P_{X,Y}$ là một phân bố xác suất trên $\mathbb{R}^2$ tập trung tại 4 điểm $A = (0,0), B = (0,1), C = (1,0), D = (1,1)$, và thoả mãn các điều kiện: $P_{X,Y}(C) + P_{X,Y}(D) = p$, $P_{X,Y}(B) + P_{X,Y}(D) = q$. Nhưng $P_{X,Y}(D)$ có thể là một số bất kỳ nằm giữa 0 và $min(p,q)$.

3.1.3 Hàm mật độ đồng thời

Định nghĩa 3.2. *Một phân bố xác suất n chiều $P_{\mathbf{X}}$ được gọi là* **liên tục tuyệt đối** *nếu nó được sinh bởi một* **hàm mật độ** $\rho_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R}_+$ *khả tích trên $\mathbb{R}^n$. Điều đó có nghĩa là với mọi miền $U \subset \mathbb{R}^n$ ta có:*

$$P_{\mathbf{X}}(U) = \int_U \rho_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n. \tag{3.7}$$

Hàm $\rho_{\mathbf{X}}$ còn được gọi là **hàm mật độ đồng thời** *của các biến ngẫu nhiên $X_1, \dots, X_n$.*

Tất nhiên, nếu một hàm ρ không âm là một hàm mật độ đồng thời trên $\mathbb{R}^n$, thì nó phải có tính chất

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \rho(x_1, \dots, x_n) dx_1 \dots dx_n = 1, \tag{3.8}$$

bởi vì xác suất của toàn bộ không gian là bằng 1. Ngược lại, mọi hàm không âm có tính chất trên là hàm mật độ của một phân bố xác suất nào đó trên $\mathbb{R}^n$.

Nếu như các phân bố xác suất P_{X_i} $(i = 1, \dots, n)$ và, $P_{\mathbf{X}}$ $(\mathbf{X} = (X_1, \dots, X_n))$ là các phân bố xác suất liên tục với các hàm mật độ

tương ứng ρ_{X_i} và $\rho_{\mathbf{X}}$, thì ta có thể viết

$$\rho_{X_1}(x) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \rho(x, y_2, \dots, y_n) dy_2 \dots dy_n, \tag{3.9}$$

và tương tự như vậy cho các hàm $\rho_{X_2}(x), \dots, \rho_{X_n}(x)$. Các hàm ρ_{X_i} được gọi là các **hàm mật độ biên** (của hàm mật độ $\rho_{\mathbf{X}}$, hay là của các biến X_i).

Ví dụ 3.3. Giả sử $\mathbf{X} = (X, Y)$ có hàm mật độ đồng thời $\rho_{\mathbf{X}}(x, y) = 1/x$ khi $0 < y \le x \le 1$ và $\rho_{\mathbf{X}}(x, y) = 0$ tại các điểm khác. Khi đó hàm mật độ biên ρ_X của X là:

$$\rho_X(x) = \int_0^x \rho_{\mathbf{X}}(x, y) dy = \int_0^x (1/x) dy = 1$$

khi $0 < x \le 1$, và $\rho_X(x) = 0$ tại các điểm khác. Điều đó có nghĩa là X có phân bố xác suất đều trên đoạn $]0, 1]$. Hàm mật độ biên của Y là:

$$\rho_Y(y) = \int_y^1 \rho_{\mathbf{X}}(x, y) dx = \int_x^1 (1/x) dx = -\ln y$$

khi $0 < y \le 1$, và $\rho_Y(y) = 0$ tại các điểm khác.

Nếu ta có hai biến ngẫu nhiên X, Y và $F : \mathbb{R}^2 \to \mathbb{R}$ là một hàm số hai biến, thì phân bố xác suất của biến ngẫu nhiên $F(X, Y)$ có thể suy ra được từ phân bố xác suất đồng thời của X và Y theo công thức:

$$\mathcal{F}_{F(X,Y)}(a) = P(F(X, Y) \le a) = P_{X,Y}(\{(x, y) \in \mathbb{R}^2 | F(x, y) \le a\}). \tag{3.10}$$

Trong trường hợp liên tục, ta cũng có thể tính được hàm mật độ của $F(X, Y)$ từ hàm mật độ đồng thời của X và Y. Ví dụ, trong trường

hợp $F(X,Y) = X + Y$, ta có công thức sau:

$$\rho_{X+Y}(z) = \int_{-\infty}^{\infty} \rho_{X,Y}(x, z-x)dx. \quad (3.11)$$

(Công thức tương tự cho các biến ngẫu nhiên rời rạc là: $P_{X+Y}(z) = \sum_x P_{X,Y}(x, z-x)$, với $P_{X,Y}(x, z-x) = P(X = x, Y = z - x)$). Tất nhiên, các khẳng định trên có thể mở rộng lên trường hợp n chiều.

Bài tập 3.3. Giả sử X là một biến ngẫu nhiên bất kỳ. Chứng minh rằng không tồn tại hàm mật độ cho vector ngẫu nhiên (X, X^3).

3.1.4 Hàm đặc trưng của vector ngẫu nhiên

Tương tự như trong trường hợp biến ngẫu nhiên 1 chiều, ta có thể định nghĩa các đại lượng đặc trưng của các vector ngẫu nhiên: nếu $F : \mathbb{R}^n \to \mathbb{C}$ là một hàm n biến bất kỳ, và $\mathbf{X} = (X_1, \ldots, X_n)$ là một vector ngẫu nhiên n, thì

$$\int_{\mathbf{x} \in \mathbb{R}^n} F(\mathbf{x}) dP_{\mathbf{X}} \quad (3.12)$$

là đại lượng đặc trưng của $\mathbf{X}$ định nghĩa bởi hàm F. Nhắc lại rằng, trong trường hợp $n = 1$ và $F(x) = x^k$, thì công thức trên cho ta moment bậc k. Trong trường hợp n tùy ý và $F(\mathbf{x}) = x_1^{k_1} \ldots x_n^{k_n}$ là một đơn thức ($\mathbf{x} = (x_1 \ldots, x_n)$), thì ta được một **moment hỗn hợp** $\mathbb{E}(X_1^{k_1} \ldots X_n^{k_n})$ của vector ngẫu nhiên n chiều. Nếu F không chỉ phụ thuộc vào x mà còn phụ thuộc vào các tham số s nào đó, thì công thức trên cho ta một hàm theo biến s, mà các giá trị của nó là các giá trị đặc trưng của $\mathbf{X}$.

Định nghĩa 3.3. Hàm đặc trưng *của vector ngẫu nhiên n chiều $\mathbf{X} = (X_1, \ldots, X_n)$ là hàm n biến $\Phi_{\mathbf{X}}(\mathbf{s})$, $\mathbf{s} = (s_1, \ldots, s_n)$, cho bởi công thức sau:*

$$\Phi_{\mathbf{X}}(s_1, \ldots, s_n) = \int_{\mathbf{x} \in \mathbb{R}^n} \exp(\sqrt{-1} \sum_{i=1}^{n} s_i x_i) dP_{\mathbf{X}}. \tag{3.13}$$

Tương tự như trong trường hợp một chiều, hàm đặv trưng trong trường hợp nhiều chiều có các tính chất sau:

Định lý 3.2. *i) Giả sử $\mathbf{X}$ là một vector ngẫu nhiên n chiều bất kỳ. Khi đó hàm đặ trưng $\Phi_{\mathbf{X}}$ của nó là một hàm liên tục đều trên $\mathbb{R}^n$, $|\Phi_{\mathbf{X}}(0)| = 1$, và $|\Phi_{\mathbf{X}}(\mathbf{s})| \leq 1$ với mọi $\mathbf{s} \in \mathbb{R}^n$.*
ii) (Công thức nghịch đảo) Giả sử $\mathbf{X}$ là một vector ngẫu nhiên n chiều, và $\mathbf{a}, \mathbf{b} \in \mathbb{R}$, sao cho $\mathbf{a} < \mathbf{b}$ và hình hộp mở n chiều $B_{\mathbf{a},\mathbf{b}} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{a} < \mathbf{x} < \mathbf{b}\}$ có xác suất của biên theo phân bố của $\mathbf{X}$ bằng 0: $P_{\mathbf{X}}(\partial B_{\mathbf{a},\mathbf{b}}) = 0$. Khi đó:

$$P(\mathbf{X} \in B_{\mathbf{a},\mathbf{b}}) =$$
$$= \frac{1}{(2\pi)^n} \lim_{T_1 \to \infty} \ldots \lim_{T_n \to \infty} \int_{-\mathbf{T} < \mathbf{s} < \mathbf{T}} \prod_{k=1}^{n} \left(\frac{e^{\sqrt{-1} s_k a_k} - e^{\sqrt{-1} s_k b_k}}{\sqrt{-1} s_k} \right) \Phi_{\mathbf{X}}(\mathbf{s}) d\mathbf{s} \tag{3.14}$$

(trong đó $d\mathbf{s} = ds_1 \ldots ds_n$ là ký hiệu độ đo Lebesgue trên $\mathbb{R}^n$).
iii) Nếu hai vector ngẫu nhiên n chiều có cùng hàm đặc trưng, thì chúng cũng có cùng phân bố xác suất trên $\mathbb{R}^n$.
iv) Giả sử $m \in \mathbb{N}$ và $\mathbb{E}(|X_i|^m) < \infty$ với mọi $i = 1, \ldots, n$. Khi đó hàm đặc trưng khả vi liên tục m lần, và

$$\mathbb{E}(X_1^{k_1} \ldots X_n^{k_n}) = \frac{1}{(\sqrt{-1})^{|\mathbf{k}|}} \frac{\partial^{|\mathbf{k}|} \Phi_{\mathbf{X}}(0)}{\partial x_1^{k_1} \ldots \partial x_n^{k_n}}, \tag{3.15}$$

với mọi $k_1, \ldots, k_n \in \mathbb{Z}_+$ *sao cho* $|\mathbf{k}| = k_1 + \ldots + k_n \leq m$.
v) (Biến đổi Fourier ngược). Trong trường hợp phân bố xác suất của $\mathbf{X}$ *là liên tục tuyệt đối với hàm mật độ* $\rho_{\mathbf{X}}$*, thì* $\Phi_{\mathbf{X}}$ *là biến đổi Fourier của* $\rho_{\mathbf{X}}$*, và* $\rho_{\mathbf{X}}$ *là biến đổi Fourier ngược của* $\Phi_{\mathbf{X}}$*:*

$$\rho_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \exp(-\sqrt{-1} \sum_k s_k x_k) \Phi_{\mathbf{X}}(\mathbf{s}) d\mathbf{s}. \tag{3.16}$$

Chứng minh của định lý trên tương tự như trường hợp 1 chiều, tuy có phức tạp hơn.

3.2 Các biến ngẫu nhiên độc lập

3.2.1 Sự độc lập của một bộ biến ngẫu nhiên

Khái niệm độc lập của các biến ngẫu nhiên là mở rộng khái niệm độc lập của các sự kiện. Về mặt triết lý, khi mà hai biến ngẫu nhiên không liên quan đến nhau, thì chúng phải độc lập với nhau. Ví dụ, "số giờ dạy học trong tuần của giáo viên" và "chiều cao của cây cau" có lẽ không liên quan gì đến nhau, có thể coi là độc lập. Nếu giả sử ta tung quân xúc sắc 3 lần, thì ta có một bộ 3 biến ngẫu nhiên, mỗi biến là kết quả của một lần tung xúc sắc. Bộ 3 biến ngẫu nhiên đó cũng có thể coi là độc lập, khi không có gì chứng tỏ kết quả của các lần tung có thể ảnh hưởng tới nhau.

Định nghĩa 3.4. *Một bộ* n *biến ngẫu nhiên* $X_1, \ldots, X_n$ *được gọi là* **độc lập** *nếu như không gian xác suất* $(\mathbb{R}^n, P_{\mathbf{X}})$ *với phân bố xác suất đồng thời* $P_{\mathbf{X}}$ *của* $X_1, \ldots, X_n$ *là tích trực tiếp của các không gian*

xác suất $(\mathbb{R}, P_{X_1}), \ldots, (\mathbb{R}, P_{X_n})$*, hay nói cách khác, với mọi tập con* $A_1, \ldots, A_n \subset \mathbb{R}$ *(nằm trong sigma-đại số Borel của* $\mathbb{R}$*) ta có*

$$P_{\mathbf{X}}(A_1 \times \ldots \times A_n) = P(X_1 \in A_1, \ldots, X_n, \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i) = \prod_{i=1}^{n} P_{X_i}(A_i). \quad (3.17)$$

Trong trường hợp các biến ngẫu nhiên $X_1, \ldots, X_n$ đều có phân bố xác suất rời rạc, điều kiện độc lập có thể được viết dưới dạng sau:

$$P_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} P_{X_i}(x_i) \quad (3.18)$$

với mọi $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

Hàm phân phối xác suất của tích trực tiếp $(\mathbb{R}, P_{X_1}) \times \ldots \times (\mathbb{R}, P_{X_n})$ chính là hàm $\mathcal{F}(x_1, \ldots x_n) = \prod_{i=1}^{n} P(X_i \leq x_i) = \prod_{i=1}^{n} \mathcal{F}_{X_i}(x_i)$. Phân bố xác suất trên $\mathbb{R}^n$ được xác định duy nhất bởi hàm phân bố xác suất của nó, bởi vậy ta có:

Định lý 3.3. *Các khẳng định sau đây là tương đương:*
i) Bộ n biến ngẫu nhiên $X_1, \ldots, X_n$ là độc lập.
ii) Hàm phân phối xác suất đồng thời $\mathcal{F}_{\mathbf{X}}$ của $X_1, \ldots, X_n$ là tích của các hàm phân phối xác suất biên $\mathcal{F}_{X_i}$:

$$\mathcal{F}_{\mathbf{X}}(x_1 \ldots, x_n) = \prod_{i=1}^{n} \mathcal{F}_{X_i}(x_i), \quad (3.19)$$

hay nói cách khác,

$$P(F_1 \leq x_1 \ldots, F_n \leq x_n) = \prod_{i=1}^{n} P(F_i \leq x_i), \quad (3.20)$$

với mọi $(x_1, \ldots, x_n) \in \mathbb{R}^n$.
iii) Hàm đặc trưng $\Phi_{\mathbf{X}}(s_1, \ldots, s_n)$ *của* $\mathbf{X} = (X_1, \ldots, X_n)$ *là tích của các hàm đặc trưng* $\Phi_{X_i}(s_i)$*:*

$$\Phi_{\mathbf{X}}(s_1, \ldots, s_n) = \prod_{i=1}^{n} \Phi_{X_i}(s_i). \tag{3.21}$$

iv) (Trường hợp liên tục tuyệt đối) Tích của các hàm mật độ biên $\rho_{X_i}(x_i)$ *của các biến* X_i *bằng hàm mật độ đồng thời:*

$$\rho_{\mathbf{X}}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \rho_{X_i}(x_i). \tag{3.22}$$

Ví dụ 3.4. Nếu A và B là hai sự kiện, và ψ_A và ψ_B là các hàm chỉ báo tương ứng của chúng ($\psi_A = 1$ nếu A xảy ra và $\psi_A = 0$ nếu A không xảy ra, và tương tự như vậy với ψ_B), thì A và B là hai sự kiện độc lập khi và chỉ khi ψ_A và ψ_B là hai biến ngẫu nhiên độc lập.

Ghi chú 3.3. Tương tự như đối với các sự kiện, có những bộ biến ngẫu nhiên không độc lập, mà trong đó các biến ngẫu nhiên độc lập với nhau theo từng đôi một. Để lấy ví dụ, ta chỉ cần lấy một bộ các sự kiện không độc lập nhưng độc lập từng đôi một, rồi lấy các hàm chỉ báo của chúng.

Bài tập 3.4. Xây dựng một ví dụ với 3 biến ngẫu nhiên X, Y, Z sao cho X độc lập với Y và Z, nhưng không độc lập với $Y + Z$.

Bài tập 3.5. Giả sử X, Y, Z là ba biến ngẫu nhiên độc lập với phân bố đều trên đoạn thẳng $]0, 1[$. Hãy tính xác suất để có thể lập được một hình tam giác với ba cạnh là X, Y, Z.

Bài tập 3.6. **Phân bố gamma** với các tham số $\alpha, \lambda > 0$ là phân bố xác suất liên tục tuyệt đối trên $\mathbb{R}$ với hàm mật độ sau: $\rho(x) =$

$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$. Ở đây $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ gọi là *hàm gamma*. Biết rằng $\Gamma(k) = (k-1)!$ với mọi $k \in \mathbb{N}$. Chứng minh (bằng qui nạp) rằng, nếu $X_1, \ldots, X_k$ là k biến ngẫu nhiên độc lập với phân bố mũ với tham số λ, thì tổng của chúng $X_1 + \ldots + X_k$ có phân bố gamma với các tham số k, λ.

Bài tập 3.7. Giả sử X và Y có phân bố xác suất đồng thời liên tục, với hàm mật độ xác suất đồng thời sau đây:

$$\rho(x,y) = \begin{cases} xe^{-x-y} \text{ khi } x, y > 0 \\ 0 \text{ tại các điểm khác} \end{cases}.$$

Hỏi rằng X và Y có độc lập với nhau không?

3.2.2 Một ví dụ không hiển nhiên về sự độc lập

Giả sử ta tung quân xúc sắc tổng cộng N lần, và mỗi lần tung thì xác suất để hiện lên mặt 1 chấm là $p = 1/6$. Gọi X là số lần tung hiện lên 1 chấm, Y là số lần tung hiện ra những mặt khác. Khi đó X và Y là hai biến ngẫu nhiên, với $X + Y = N$. Nếu số lần tung N là một số cố định, thì X và Y không độc lập với nhau, vì $P(X = a, Y = b) = 0$ nếu $a + b \neq N$.

Bây giờ ta giả sử rằng bản thân tổng số lần tung N là một số ngẫu nhiên tuân theo luật Poisson với tham số λ:

$$P(N = n) = e^{-\lambda} \frac{\lambda^n}{n!},\ n \in \mathbb{Z}_+$$

Khi đó $P(X = a, Y = b) \neq 0$ với mọi $a, b \in \mathbb{Z}_+$. Xác suất có điều kiện $P(X = x|N = n)$ tuân theo phân bố nhị phân:

$$P(X = x|N = n) = C_n^x p^x (1-p)^{n-x}$$

Từ đó suy ra:

$$
\begin{aligned}
P(X = x, Y = y) &= P(X = x | N = x + y).P(N = x + y) \\
&= C_{x+y}^{x} p^x (1-p)^y \frac{\lambda^{x+y}}{(x+y)!} e^{-\lambda} = \frac{(\lambda p)^x}{x!} \frac{(\lambda(1-p))^y}{y!} e^{-\lambda} v, \\
P(X = x) &= \textstyle\sum_{y \in \mathbb{Z}_+} P(X = x, Y = y) = \frac{(\lambda p)^x}{x!}.C_1, \\
P(Y = y) &= \textstyle\sum_{x \in \mathbb{Z}_+} P(X = x, Y = y) = \frac{(\lambda(1-p))^y}{y!}.C_2,
\end{aligned}
$$

với $C_1 = e^{-\lambda p}, C_2 = e^{-\lambda(1-p)}$, và $P(X = x).P(y = y) = P(X = x, Y = y)$. Điều đó có nghĩ là, trong trường hợp này (khi mà tổng $N = X + Y$ tuân theo phân bố Poisson), hai biến X và Y độc lập với nhau!

3.2.3 Một số hệ quả của sự độc lập

Định lý 3.4. *Giả sử $X_1, \ldots, X_n$ là một bộ n biến ngẫu nhiên độc lập, và $g_1, \ldots, g_n$ là các hàm số thực. Khi đó các biến ngẫu nhiên $g_1(X_1), \ldots, g_n(X_n)$ cũng độc lập với nhau.*

Chứng minh. Với các tập con $A_1, \ldots, A_n \subset \mathbb{R}$ bất kỳ, ta có: $P(g_1(X) \in A_1, \ldots, g_n(X) \in A_n) = P(X_1 \in g_1^{-1}(A_1), \ldots, X_1 \in g_n^{-1}(A_n)) = \prod_i P(X_i \in g_i^{-1}(A_i)) = \prod_i P(g(X_i) \in A_i)$, và do đó các biến ngẫu nhiên $g_1(X_1), \ldots, g_n(X_n)$ độc lập với nhau. □

Tương tự như vậy, ta có mệnh đề sau (chứng minh của nó là bài tập dành cho bạn đọc):

Mệnh đề 3.5. *Nếu X_1, X_2, X_3 là một bộ 3 biến ngẫu nhiên độc lập, và $\phi : \mathbb{R}^2 \to \mathbb{R}$ là một hàm hai biến, thì các biến ngẫu nhiên $\phi(X_1, X_2)$ và X_3 độc lập với nhau.*

Nhắc lại rằng, nếu X là một biến ngẫu nhiên, thì hàm sinh xác suất của nó là hàm $G_X(z) = \mathbb{E}(z^X)$, và hàm đặc trưng của nó là hàm $\Phi_X(s) = \mathbb{E}(\exp(\sqrt{-1}sX))$.

Định lý 3.6. *Nếu $X_1, \ldots, X_n$ là một bộ n biến ngẫu nhiên độc lập, thì:*
i)

$$\mathbb{E}(\prod_i X_i) = \prod_i \mathbb{E}(X_i), \tag{3.23}$$

ii)

$$var(\sum_i X_i) = \sum_i var(X_i), \tag{3.24}$$

iii)

$$G_{\sum_i X_i}(z) = \prod_i G_{X_i}(z), \tag{3.25}$$

iv)

$$\Phi_{\sum_i X_i}(s) = \prod_i \Phi_{X_i}(s). \tag{3.26}$$

Chứng minh. i) và ii). Chúng ta sẽ chứng minh cho trường hợp $n = 2$, và hai biến ngẫu nhiên X_1, X_2 chỉ nhận một số hữu hạn các giá trị $\{a_1, \ldots, a_k\}$ và $\{b_1, \ldots, b_m\}$ tương ứng. Khi đó ta có:

i) $\mathbb{E}(X_1X_2) = \sum_{ij} a_ib_jP(X_1 = a_i, X_2 = b_j)$
$= \sum_{ij} a_ib_jP(X_1 = a_i)P(X_2 = b_j)$
$= (\sum_i a_iP(X_1 = a_i))(\sum_j b_jP(X_2 = b_j)) = \mathbb{E}(X_1)\mathbb{E}(X_2)$.

ii) $var(X_1 + X_2) = \mathbb{E}((X_1 + X_2)^2) - \mathbb{E}(X_1 + X_2)^2 =$
$= \mathbb{E}(X_1^2) + 2\mathbb{E}(X_1X_2) + \mathbb{E}(X_2^2) - (\mathbb{E}(X_1) + \mathbb{E}(X_2))^2 =$
$= \mathbb{E}(X_1^2) + 2\mathbb{E}(X_1X_2) + \mathbb{E}(X_2^2) - \mathbb{E}(X_1)^2 - 2\mathbb{E}(X_1)\mathbb{E}(X_2) - \mathbb{E}(X_2)^2 =$
$= \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 + \mathbb{E}(X_2^2) - \mathbb{E}(X_2)^2 = var(X_1) + var(X_2)$.

Trường hợp các biến ngẫu nhiên nhận vô hạn các giá trị có thể suy ra từ trường hợp với hữu hạn các giá trị bằng cách lấy giới hạn. Trường hợp n tùy ý suy ra từ trường hợp $n = 2$ bằng qui nạp.

iii) Ta có $G_{\sum_i X_i}(z) = \mathbb{E}(z^{\sum X_i}) = \mathbb{E}(\prod z^{X_i}) = \prod_i \mathbb{E}(z^{X_i}) = \prod_i G_{X_i}(z)$, do các biến z^{X_i} độc lập với nhau.

iv) Chứng minh hoàn toàn tương tự. □

Ví dụ 3.5. Giả sử X và Y là hai biến ngẫu nhiên độc lập có phân bố Poisson với các tham số là λ và γ tương ứng. Khi đó $X + Y$ cũng có phân bố Poisson với tham số là $\lambda + \gamma$. Để thấy điều đó, ta có thể tính $P(X + Y = k)$ qua công thức

$$P(X+Y = k) = \sum_h P(X = h, Y = k-h) = \sum_h P(X = h)P(Y = k-h),$$

hoặc là ta có thể lý luận như sau: Hàm sinh của X là $G_X(z) = \exp(\lambda(z-1))$, của Y là $G_Y(z) = \exp(\gamma(z-1))$. Vì X và Y độc lập với nhau nên hàm sinh của $X + Y$ là $G_{X+Y}(z) = G_X(z)G_Y(z) = \exp(\lambda(z-1))\exp(\gamma(z-1)) = \exp((\lambda+\gamma)(z-1))$ là hàm sinh của phân bố Poisson với tham số $\lambda + \gamma$. Bởi vậy $X + Y$ có phân bố này.

Bài tập 3.8. Giả sử X và Y là hai biến ngẫu nhiên độc lập tuân theo các phân bố normal $\mathcal{N}(\mu_1, \sigma_1^2)$ và $\mathcal{N}(\mu_2, \sigma_2^2)$ tương ứng. Hãy tính hàm đặc trưng của $X + Y$, và từ đó suy ra rằng $X + Y$ tuân theo phân bố normal $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Bài tập 3.9. Tung một quân xúc sắc nhiều lần, cho đến khi tổng của tất các các số hiện lên trong các lần tung đạt ít nhất 1000. Gọi γ là xác suất để số lần phải tung lớn hơn 350. Dùng bất đẳng thức Chebyschev để tìm một đánh giá chặn trên của γ.

Bài tập 3.10. Tung một con xúc sắc 5 lần. Dùng hàm sinh xác suất, hãy tính xác suất để tổng các số hiện lên trong 5 lần tung là 15.

Bài tập 3.11. Giả sử $X_1, X_2, X_3, \ldots$ là một dãy các biến ngẫu nhiên độc lập có cùng một phân bố xác suất, với kỳ vọng $\mu < 0$ và phương sai $\sigma^2 < \infty$. Gọi $S_n = X_1 + \ldots + X_n$ là tổng của n biến ngẫu nhiên đầu tiên. Dùng bất đẳng thức Chebyschev để chứng minh rằng, với mọi $c \in \mathbb{R}$, ta có $\lim_{n\to\infty} P(S_n \geq c) = 0$.

3.3 Luật số lớn

3.3.1 Dạng yếu của luật số lớn cho phân bố bất kỳ

Giả sử $X_1, X_2, X_3, \ldots$ là một dãy vô hạn các biến ngẫu nhiên độc lập có cùng một phân bố xác suất với kỳ vọng μ và phương sai σ^2 hữu hạn. Đặt $S_n = X_1 + \ldots + X_n$. Ta có mở rộng sau đây của định lý 1.4:

Định lý 3.7 (Luật số lớn). *Với mọi $\epsilon > 0$ ta có*

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right) = 1. \tag{3.27}$$

Định lý Bernoulli 1.4 là trường hợp riêng của định lý trên, khi mà X_i chỉ nhận hai giá trị 0 và 1.

Chứng minh. Do các biến ngẫu nhiên X_i độc lập với nhau nên $var(S_n) = \sum_{i=1}^n var(X_i) = n\sigma^2$, và $\mathbb{E}(S_n) = \sum_{i=1}^n \mathbb{E}(X_i) = n\mu$. Áp

dụng bất đẳng thức Chebyschev, ta có:

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = P(|S_n - \mathbb{E}(S_n)| \geq n\epsilon) \leq \frac{var(S_n)}{(n\epsilon)^2} = \frac{\sigma^2}{n\epsilon^2} \stackrel{n\to\infty}{\longrightarrow} 0, \tag{3.28}$$

từ đó suy ra điều phải chứng minh. □

Ghi chú 3.4. Chứng minh của định lý 3.7 tất nhiên có thể dùng được cho định lý 1.4, và nó đơn giản hơn cách chứng minh định lý 1.4 viết phía trên. Tuy nhiên, cách chứng minh định lý 1.4 phía trên cho một đánh giá hội tụ tốt hơn: dãy số dạng a^n, với $0 < a < 1$, hộ tụ về 0 nhanh hơn là dãy số $1/n$.

Ghi chú 3.5. Kể cả trong trường hợp với kỳ vọng μ hữu hạn nhưng phương sai vô hạn, định lý 3.7 vẫn đúng, nhưng chứng minh phức tạp hơn, và khi đó nó được gọi là **định lý Khinchin**. (Có thể xem, chẳng hạn, [5] và [6]).

Bài tập 3.12. Một sòng bạc hợp pháp được rao bán, và bạn là nhà đầu tư muốn mua nó. Nhưng trước khi mua nó bạn muốn biết lợi nhuận hàng năm của nó bao nhiêu. Sòng bạc này chỉ chuyên về trò quay vòng đỏ đen. Mỗi bàn quay có 37 ô: 18 ô đỏ, 18 ô đen, và 1 ô nhà cái. Nếu khi quay vòng kim chỉ vào ô cùng mầu với ô đặt cọc thì người chơi thắng, đặt 1 ăn 1, còn nếu kim chỉ vào ô khác màu hoặc vào ô nhà cái thì người chơi mất tiền đặt cọc. Nói cách khác, cứ mỗi lần đặc cọc, thì xác suất để nhà cái thắng số tiền đặt cọc đó là 19/37, và để nhà cái thua số tiền đặt cọc đó là 18/37. Biết rằng trong năm sòng bạc mở cửa cả 365 ngày, mỗi ngày trung bình các người chơi đặt cọc tổng cộng 50 nghìn euro. Giải thích tại sao luật số lớn lại có thể dùng để tính ước lượng số tiền thu về được trong 1 năm của sòng

bạc từ trò chơi quay vòng đỏ đen (trước khi trừ chi phí hoạt động), và hãy tính con số này.

3.3.2 Dạng mạnh của luật số lớn

Định lý 3.8 (Luật số lớn). *Giả sử $X_1, X_2, X_3, \ldots$ là một dãy vô hạn các biến độc lập có cùng một phân bố xác suất với kỳ vọng bằng μ và $\mathbb{E}(X_1^4)$ hữu hạn. Khi đó ta có:*

$$P\left(\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{n} = \mu\right) = 1, \tag{3.29}$$

hay nói cách khác, hầu như chắc chắn rằng, $\frac{\sum_{i=1}^{n} X_i}{n}$ tiến tới μ khi n tiến tới vô cùng.

Ghi chú 3.6. Tất nhiên, dạng mạnh của luật số lớn mạnh hơn dạng yếu, và bởi vậy cũng đòi hỏi điều kiện mạnh hơn: nếu một phân bố xác suất nào đó thỏa mãn luật số lớn mạnh, thì nó cũng nghiễm nhiên thỏa mãn luật số lớn yếu, tuy điều ngược lại không đúng.

Trước khi chứng minh định lý 3.8, ta cần hiểu chính xác ý nghĩa toán học của định lý trên, và cần có mô hình xác suất tự nhiên cho sự kiện $\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{n} = \mu$, tức là mô tả tập hợp tất cả các tình huống xảy ra sự kiện này như là một tập con trong một không gian xác suất nào đó. Không gian xác suất ở đây sẽ tích trực tiếp của một dãy vô hạn các không gian xác suất.

3.3.3 Tích của một dãy vô hạn các không gian xác suất

Giả sử ta có một dãy vô hạn các không gian xác suất $(\Omega_i, P_i),\ i = 1, 2, 3, \ldots$. Khi đó **tích trực tiếp** của chúng là không gian

$$\Omega = \prod_{i=1}^{\infty} \Omega_i = \{(x_1, x_2, x_3, \ldots) | x_i \in \Omega_i \ \forall i \in \mathbb{N}\}. \tag{3.30}$$

Mỗi phần tử của Ω là một dãy $\mathbf{x} = (x_i)_{i\in\mathbb{N}}$ các phần tử x_i: $x_i \in \Omega_i$ với mọi $i \in \mathbb{N}$. **Phân bố xác suất tích** trên Ω được cho bởi công thức sau: nếu $A_i \subset \Omega_i$ sao cho tồn tại $P_i(A_i)$ với mọi $i \in \mathbb{N}$, thì theo định nghĩa,

$$P(\prod_{i=1}^{\infty} A_i) := \prod_{i=1}^{\infty} P_i(A_i) := \lim_{n\to\infty} \prod_{i=1}^{n} P_i(A_i). \tag{3.31}$$

Ở đây $\prod_{i=1}^{\infty} A_i = \{(x_1, x_2, x_3, \ldots) | x_i \in A_i \ \forall i \in \mathbb{N}\}$ là một tập con của Ω có dạng tích trực tiếp. Chú ý rằng tích vô hạn

$$\prod_{i=1}^{\infty} P_i(A_i) = \lim_{n\to\infty} \prod_{i=1}^{n} P_i(A_i)$$

tồn tại và không âm, bởi vì dãy số $(\prod_{i=1}^{n} P_i(A_i))_{n\in\mathbb{N}}$ là một dãy đơn điệu không tăng không âm. Tích này có thể bằng 0 kể cả khi $P_i(A_i) > 0$ với mọi i. Thế nhưng nếu $P_i(A_i) > 0$ với mọi i, và $P_i(A_i) = 1$ với hầu hết mọi i trừ một số hữu hạn các giá trị của i, thì tích này có thể coi như là một tích hữu hạn, với giá trị dương.

Đối với các tập con của Ω không có dạng tích trực tiếp, thì xác suất của chúng có thể tính được từ xác suất của các tập con có dạng tích trực tiếp, thông qua các tiên đề của xác suất. Ví dụ, xác suất của

hợp của hai tập con có dạng tích trực tiếp là:

$$\begin{aligned} P(\prod_i A_i \cup \prod_i B_i) &= P(\prod_i A_i) + P(\prod_i B_i) - P((\prod_i A_i) \cap (\prod_i B_i)) \\ &= P(\prod_i A_i) + P(\prod_i B_i) - P(\prod_i (A_i \cap B_i)) \\ &= \prod_i P_i(A_i) + \prod_i P_i(B_i) - \prod_i P_i(A_i \cap B_i). \end{aligned}$$

Bằng cách đó, ta có thể định nghĩa xác suất của mọi tập con của Ω mà nằm trong sigma-đại số sinh bởi các tập con có dạng tích trực tiếp (qua các phép: phần bù, giao, hợp, và hợp một dãy vô hạn). Độ đo xác suất P trên Ω định nghĩa như trên, cùng với sigma-đại số này, được gọi là **tích trực tiếp** của các độ đo xác suất P_i.

Nếu ta có một dãy vô hạn các biến ngẫu nhiên độc lập X_i,, thì ta có thể coi nó như một vector vô hạn chiều $(X_i)_{i\in\mathbb{N}}$, và vector vô hạn chiều này sinh ra trên không gian vô hạn chiều $\mathbb{R}^{\mathbb{N}}$ một phân bố xác suất P, chính là tích trực tiếp của các phân bố xác suất P_i của X_i trên $\mathbb{R}$:

$$(\mathbb{R}^{\mathbb{N}}, P) = \prod_{i=1}^{\infty} (\mathbb{R}, P_i) \qquad (3.32)$$

Trong trường hợp các biến X_i có cùng phân bố xác suất, tức là $P_i = P_1$ với mọi $i \in \mathbb{N}$, ta có thể viết

$$(\mathbb{R}^{\mathbb{N}}, P) = (\mathbb{R}, P_1)^{\mathbb{N}} \qquad (3.33)$$

Tích vô hạn $(\mathbb{R}, P_i)^{\mathbb{N}}$ này có thể dùng làm mô hình không gian xác suất trong định lý 3.8. Khi đó sự kiện $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu$ ứng với tập

hợp con $\{(x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} | \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\}$ trong $\mathbb{R}^{\mathbb{N}}$. Định lý 3.8 tương đương với khẳng định

$$P(A) = 0, \tag{3.34}$$

trong đó $A = \mathbb{R}^{\mathbb{N}} \setminus \{(x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} | \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\}$.

Có một chi tiết kỹ thuật là: tập A trên không nhất thiết nằm trong sigma-đại số của $(\mathbb{R}, P_1)^{\mathbb{N}}$ xây dựng như trên. Khi đó ta phải hiểu $P(A) = 0$ như thế nào? Vấn đề này dẫn đến định nghĩa sau:

Định nghĩa 3.5. *Giả sử B là một tập đo được của một không gian xác suất (Ω, P), với $P(B) = 0$, và A là tập con của B. Khi đó A sẽ được gọi là* **tập con có thể bỏ qua**, *và ta cũng viết*

$$P(A) = 0.$$

Nói cách khác, nếu một tập có xác suất bằng 0, thì ta coi mọi tập con của nó cũng có xác suất bằng 0, kể cả khi các tập con đó không nằm trong sigma-đại số ban đầu. Ta có thể mở rộng sigma-đại số để chứa tất cả các tập con như vậy.

3.3.4 Chứng minh định lý 3.8

Chúng ta sẽ chia chứng minh của định lý 3.8 thành một số bước, mỗi bước chúng ta sẽ viết dưới dạng một bổ đề.

Bổ đề 3.9 (Tiêu chuẩn xác suất bằng 0). *Nếu tồn tại một dãy các tập con $A \subset A_n \subset (\Omega, P)$ sao cho A_n đo được và $\lim_{n\to\infty} P(A_n) = 0$, thì $P(A) = 0$.*

Chứng minh của bổ đề trên là bài tập dành cho bạn đọc.

Giả sử $X_1, X_2, X_3, \ldots$ là một dãy vô hạn các biến độc lập có cùng một phân bố xác suất với kỳ vọng bằng μ và $\mathbb{E}(X_1^4)$ hữu hạn, như trong định lý 3.8. Đặt $Y_i = X_i - \mu$ (để chuyển định lý về trường hợp với kỳ vọng bằng 0). Dễ thấy rằng, điều kiện $\mathbb{E}(X_1^4)$ hữu hạn tương đương với điều kiện moment trung tâm bậc 4 $\mu_4 = \mathbb{E}(Y_1^4)$ hữu hạn, và suy ra điều kiện phương sai $\sigma^2 = \mathbb{E}(Y_1^2)$ hữu hạn.

Bổ đề 3.10. *Với mọi $n \in \mathbb{N}$ ta có*

$$\mathbb{E}((\sum_{i=1}^{n} Y_i)^4) = n\mu_4 + 3n(n-1)\sigma^4 \leq Cn^2, \tag{3.35}$$

trong đó $C = 3\sigma^4 + \mu_4$ là một hằng số (không phụ thuộc vào n).

Chứng minh. Ta có

$$(\sum_{i=1}^{n} Y_i)^4 = \sum_{i=1}^{n} Y_i^4 + 6\sum_{i<j} Y_i^2 Y_j^2 + 4\sum_{i\neq j} Y_i^3 Y_j + 6\sum_{i\neq j\neq k} Y_i^2 Y_j Y_k + \sum_{i\neq j\neq k\neq l} Y_i Y_j Y_k Y_l.$$

Do các biến ngẫu nhiên $Y_1, \ldots, Y_n$ độc lập với nhau và có kỳ vọng bằng 0 nên $\mathbb{E}(Y_i^3 Y_j) = \mathbb{E}(Y_i^3)\mathbb{E}(Y_j) = 0$ với mọi $i \neq j$, và tương tự như vậy, $\mathbb{E}(Y_i^2 Y_j Y_k) = \mathbb{E}(Y_i Y_j Y_k Y_l) = 0$ với mọi $i \neq j \neq k \neq l$. Bởi vậy

$$\mathbb{E}((\sum_{i=1}^{n} Y_i)^4) = \sum_{i=1}^{n} \mathbb{E}(Y_i^4) + 6\sum_{i<j} \mathbb{E}(Y_i^2)\mathbb{E}(Y_j^2) = n\mu_4 + 3n(n-1)\sigma^4.$$

Bổ đề 3.11. *Với mọi $k \in \mathbb{N}$ tồn tại một hằng số $C_k = k^4 C$ sao cho*

$$P(|\frac{1}{n}\sum_{i=1}^{n} Y_i| \geq \frac{1}{k}) < \frac{C_k}{n^2} \tag{3.36}$$

với mọi $n \in \mathbb{N}$.

Bổ đề 3.11 suy ra trực tiếp từ bổ đề 3.10 và bất đẳng thức Markov.

Bổ đề 3.12. *Với mọi $k \in \mathbb{N}$ tồn tại một số $m_k \in \mathbb{N}$ sao cho, đặt*

$$B_k = \bigcup_{n \geq m_k} (|\frac{1}{n}\sum_{i=1}^{n} Y_i| \geq \frac{1}{k}), \tag{3.37}$$

ta có

$$P\left(B_k\right) < \frac{1}{2^k}. \tag{3.38}$$

Bổ đề 3.12 suy ra trực tiếp từ bổ đề 3.11 và sự hội tụ của chuỗi số $\sum_{n=1}^{\infty} \frac{1}{n^2}$.

Nhắc lại rằng, A là sự kiện "$\frac{1}{n}\sum_{i=1}^{n} Y_i$ không tiến tới 0 khi n tiến tới vô cùng", và định lý 3.8 tương đương với khẳng định $P(A) = 0$. Đặt

$$A_k = \bigcup_{h > k} B_h. \tag{3.39}$$

Bổ đề 3.13. *Ta có*

$$A \subset A_k \text{ và } P(A_k) < \frac{1}{2^k} \tag{3.40}$$

với mọi $k \in \mathbb{N}$.

Khẳng định $A \subset A_k$ suy ra trực tiếp từ định nghĩa về giới hạn, còn bất đẳng thức $P(A_k) < 1/2^k$ là hệ quả trực tiếp của bổ đề 3.12. Từ bổ đề cuối cùng ta suy ra $P(A) = 0$, là điều cần phải chứng minh. □

3.4 Sự tương quan giữa các biến ngẫu nhiên

3.4.1 Hiệp phương sai

Định nghĩa 3.6. *Nếu X, Y là hai biến ngẫu nhiên, thì* **hiệp phương sai** *(covariance) của chúng là đại lượng*

$$cov(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))). \quad (3.41)$$

Trong trường hợp đặc biệt, khi $X = Y$, từ định nghĩa trên ta có khẳng định sau: hiệp phương sai của một biến ngẫu nhiên với chính nó chính là phương sai của nó:

$$cov(X, X) = \mathbb{E}((X - \mathbb{E}(X))^2) = var(X) = \sigma(X)^2. \quad (3.42)$$

Ý nghĩa của hiệp phương sai $cov(X, Y)$ như sau: nó đo độ dao động "cùng hướng" hay "ngược hướng" của X và Y. Ở đây ta hình dung là X và Y dao động quanh trung điểm (giá trị kỳ vọng) tương ứng của chúng. Nếu như X và Y luôn dao động cùng hướng, tức là X dao động lên trên trung điểm ($X - \mathbb{E}(X) > 0$) mỗi khi Y cũng dao động lên trên trung điểm, và X dao động xuống dưới mỗi khi Y cũng dao động xuống dưới, thì $(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))$ luôn có có giá trị lớn hơn hoặc bằng 0, và $cov(X, Y)$ là số dương. Ngược lại, nếu X và

Y dao động ngược hướng, thì $cov(X, Y)$ là số âm. Trong trường hợp chung, $cov(X, Y)$ là số âm hay số dương tùy thuộc vào việc X và Y dao động ngược hướng nhiều hơn hay là dao động cùng hướng nhiều hơn.

Định lý 3.14. *i) Một công thức khác để tính hiệp phương sai là:*

$$cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \tag{3.43}$$

ii) Nếu hai biến X và Y độc lập với nhau, thì $cov(X, Y) = 0$.

Chứng minh. i) Ta có:

$$\mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))) = \mathbb{E}(XY-\mathbb{E}(X)Y-X\mathbb{E}(Y)+\mathbb{E}(X)\mathbb{E}(Y))$$
$$= \mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)-\mathbb{E}(X)\mathbb{E}(Y)+\mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y).$$

ii) Khi X và Y độc lập với nhau thì $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, do đó $cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$ □

Định lý 3.15. *Hiệp phương sai có các tính chất sau:*
i) Đối xứng: $cov(X, Y) = cov(X, Y)$
ii) Tuyến tính: $cov(a_1.X_1 + a_2.X_2, Y) = a_1.cov(X_1, X) + a_2.cov(X_2, Y)$
iii) Bất biến theo xê dịch: $cov(X + a, Y) = cov(X, Y)$

Các tính chất trên suy ra trực tiếp từ định nghĩa. Tính chất tuyến tính cũng đúng với biến Y, nên ta nói rằng $cov(X, Y)$ có tính chất **song tuyến tính**.

3.4.2 Hệ số tương quan

Do tính chất song tuyến tính của hiệp phương sai $cov(X, Y)$, ta có thể chia $cov(X, Y)$ cho $\sigma(X)\sigma(Y)$ để được một đại lượng có bậc

thuần nhất bằng 0, tức là không thay đổi khi ta nhân X, Y với các hằng số. Đại lượng đó được gọi là hệ số tương quan (correlation) của X và Y:

Định nghĩa 3.7. *Nếu hai biến ngẫu nhiên X, Y có độ lệch chuẩn $\sigma(X), \sigma(Y)$ khác 0, thì* **hệ số tương quan** *của chúng là đại lượng sau:*

$$r(F, G) = \frac{cov(F, G)}{\sigma(F)\sigma(G)}. \tag{3.44}$$

Định lý 3.16. *Nếu X, Y là hai biến ngẫu nhiên có độ lệch chuẩn khác 0, thì ta luôn có*

$$-1 \leq r(X, Y) \leq 1. \tag{3.45}$$

Hơn nữa, $r(X, Y) = 1$ khi và chỉ khi X, Y có quan hệ tuyến tính với nhau với hệ số dương, có nghĩa là tồn tại một số thực dương $a > 0$ và một số thực b sao cho $X = aY + b$ hầu khắp mọi nơi. Ngược lại, $r(X, Y) = -1$ khi và chỉ khi X, Y có quan hệ tuyến tính với nhau với hệ số âm, có nghĩa là tồn tại một số thực dương $a < 0$ và một số thực b sao cho $X = aY + b$ hầu khắp mọi nơi.

Định lý trên là hệ quả trực tiếp của bất đẳng thức Cauchy-Schwarz sau:

Định lý 3.17. *(Bất đẳng thức Cauchy-Schwarz). Nếu U, V là hai biến ngẫu nhiên thực bất kỳ thì ta luôn có*

$$\mathbb{E}(UV)^2 \leq \mathbb{E}(U^2)\mathbb{E}(V^2). \tag{3.46}$$

Dấu bằng xảy ra khi và chỉ khi U và V tỷ lệ thuận với nhau, tức là hoặc là $V = 0$ hầu khắp mọi nơi hoặc là ta có thể viết $U = cV$ hầu khắp mọi nơi, với c là một hằng số.

Trường hợp mà không gian xác suất là hữu hạn với phân bố xác suất đều, bất đẳng thức Cauchy-Schwarz có dạng cổ điển quen thuộc sau: Với các số thực a_i, b_i, $i = 1, \ldots, n$, bất kỳ, ta có:

$$(\sum_i a_i b_i)^2 \leq (\sum_i a_i^2).(\sum_i b_i)^2. \qquad (3.47)$$

Để chứng minh bất đẳng thức cổ điển trên, chỉ cần kiểm tra rằng

$$(\sum_i a_i b_i)^2 - (\sum a_i^2).(\sum_i b_i)^2 = -\sum_{i<j}(a_i b_j - a_j b_i)^2 \leq 0. \qquad (3.48)$$

Dấu bằng xảy ra khi và chỉ khi $a_i b_j = a_j b_i$ với mọi i, j, có nghĩa là dãy số (a_i) tỷ lệ thuận với dãy số (b_i). Trường hợp tổng quát của bất đẳng thức Cauchy-Schwarz trên không gian xác suất chẳng qua là giới hạn của trường hợp cổ điển quen thuộc trên.

Chứng minh bất đẳng thức Cauchy-Schwarz trong trường hợp tổng quát: Ta có thể viết $U = U_1 + aV$ với $a = \mathbb{E}(UV)/\mathbb{E}(V^2)$. Khi đó ta có $\mathbb{E}(UV)^2 = a^2\mathbb{E}(V^2)^2$, $\mathbb{E}(U_1.V) = 0$, $\mathbb{E}(U^2) = \mathbb{E}(U_1^2 + 2a.U_1.V + a^2.V^2) = \mathbb{E}(U_1^2) + a^2.\mathbb{E}(V^2) \geq a^2.\mathbb{E}(V^2)$, và bởi vậy $\mathbb{E}(U^2).\mathbb{E}(V^2) \geq a^2.\mathbb{E}(V^2).\mathbb{E}(V^2) = \mathbb{E}(UV)^2$. □

Trong bất đẳng thức Cauchy-Schwarz, nếu ta đặt $U = F - \mathbb{E}(F)$ và $V = G - \mathbb{E}(G)$ thì ta được bất đẳng thức $cov(F, G)^2 \leq \sigma(F)^2.\sigma(G)^2$, từ đó suy ra $-1 \leq r(F, G) \leq 1$.

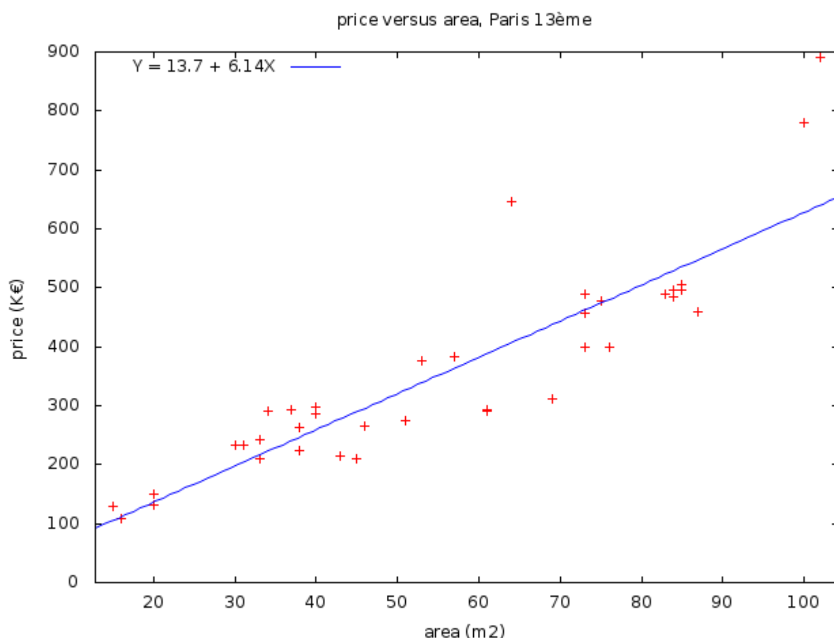
Ghi chú 3.7. Đại lượng $\mathbb{E}(UV)$ được gọi là **tích vô hướng** của U và V. Với tích vô hướng này, không gian các biến ngẫu nhiên (trên một không gian xác suất cố định nào đó) trở thành **không gian tiền Hilbert** (pre-Hilbert space).

Ví dụ 3.6. (*Giá nhà và diện tích nhà*). Dãy số liệu sau về giá rao bán các căn hộ ở Quận 13 thành phố Paris được lấy từ một trang web bất động sản vào ngày 12/10/2009. Để làm ví dụ ở đây, chúng ta sẽ chỉ lấy 40 quảng cáo xuất hiện mới nhất, trong số hàng nghìn quảng cáo:

(102, 890), (45, 210), (40, 286), (53, 375), (69, 311), (64, 645), (84,498), (38, 262), (33, 210), (38, 223), (33, 242), (15, 129), (73, 456), (31, 233), (16, 109), (40, 297), (85, 495), (84, 485), (100, 780), (83, 490), (87, 460), (51, 275), (40, 297), (85, 495), (85, 505), (43, 215), (46, 265), (75, 477), (61, 293), (76, 399), (73, 399), (73, 490), (85, 495), (37, 292), (34, 290), (30, 232), (20, 150), (57, 383), (20, 132), (61, 290)

Trong dãy số liệu trên, mỗi cặp số gồm 2 số: số thứ nhất là diện tích của căn hộ, tính theo đơn vị m2, số thứ hai là giá rao bán, tính theo đơn vị nghìn euro. Ví dụ, (102, 890) có nghĩa là một căn hộ rộng 102m2 được rao bán với giá 890 nghìn Euro. Chúng ta sẽ coi không gian xác suất ở đây gồm 40 phần tử, với phân bố xác suất đều, mỗi phần tử ứng với một căn hộ được rao bán trong 40 quảng cáo phía trên. (Không gian xác suất này được gọi là **không gian xác suất thực nghiệm**).

Từ số liệu trên, ta có thể tính ra hệ số tương quan giữa biến $X =$ "diện tích của căn hộ ở Quận 13" và biến $Y =$ "giá rao bán căn hộ ở Quận 13" (tại thời điểm 12/10/2009) bằng 0,888. Con số này có thể tính được bằng tay, nhưng cũng có thể dùng phần mềm máy tính để tính, sẽ nhanh hơn. Đặc biệt là khi bảng số liệu rất lớn (không gian xác suất có rất nhiều phần tử), thì cách tính tốt nhất là nhập số liệu vào máy rồi tính bằng máy. Để tính hệ số tương quan trong ví dụ



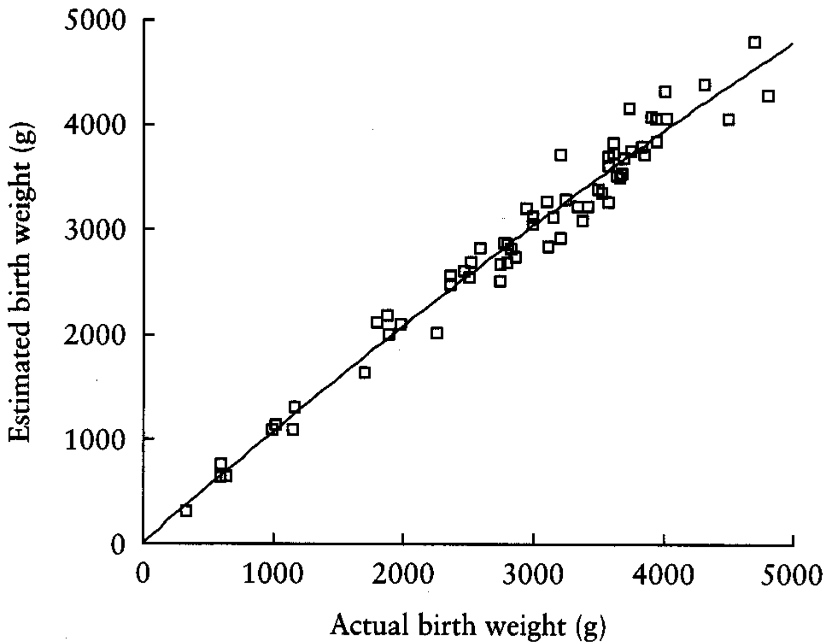
Hình 3.1: Diện tích căn hộ và giá rao bán tại Quận 13, Paris, tháng 10/2009

này, các tác giả dùng phần mềm **gretl** (viết tắt của: Gnu Regression, Econometrics and Time-series Library), là một phần mềm nhỏ tự do mã mở, có giao diện trực giác dễ sử dụng.

Hình 3.1, do chương trình gretl vẽ ra, là **đồ thị phân tán** (scatterplot) của hai biến "diện tích căn hộ" và "giá căn hộ" trong ví dụ trên.

Ví dụ 3.7. Trọng lượng trẻ em lúc sinh ra, và ước lượng bằng soi siêu âm. Một nghiên cứu thống kê trong y học của Schild, Fimmers và

Hansmann[2] trên 65 trẻ em cho thấy phương pháp ước lượng trọng lượng trẻ em trước lúc sinh ra bằng soi siêu âm 3 chiều cho kết quả rất tốt: hệ số tương quan giữa ước lượng và trọng lượng thực tế lúc sinh ra là 0,976. Xem đồ thị phân tán trên hình 3.2.



Hình 3.2: Trọng lượng ước lượng bằng soi siêu âm và trọng lượng thực tế

[2] R.L. Schild, R. Fimmers, L. Hansmann, Fetal weight estimation by three-dimensional ultrasound, Ultrasound in Obstetrics and Gynecology, 16 (2000), 445–452.

Bài tập 3.13. Xây dựng một ví dụ với hai biến ngẫu nhiên không độc lập với nhau, nhưng có hiệp phương sai bằng 0.

Bài tập 3.14. (*Giá xe ô tô và tuổi của xe*). Dãy số liệu sau về tuổi của xe Mercedes C220 cũ (số năm mà xe đã chạy) và giá rao bán xe (tính theo euro) được lấy từ trang web vivastreet (chuyên về quảng cáo bán đồ cũ) ngày 25/10/2009: (13, 3000), (4, 17500), (7, 9900), (3, 17800), (6,11500), (6, 14000), (4,18000), (6, 15000), (10, 5490), (8, 12000), (1, 32500), (10, 6500), (9, 5900), (3, 24200), (11, 6000), (2, 21000), (9,10700), (0, 30000), (8, 9800), (13, 4200). Hãy tính hệ số tương quan giữa hai biến "tuổi của xe" và "giá rao bán xe" cho các xe Mercedes C220 cũ, dựa theo dãy số liệu trên.

Bài tập 3.15. Tìm trọng lượng và chiều cao của một nhóm người (ví dụ một lớp học), rồi tính hệ số tương quan giữa hai biến "trọng lượng" và "chiều cao" của những người trong nhóm đó.

3.4.3 Quan hệ tuyến tính với sai số bình phương nhỏ nhất

Nhắc lại rằng, nếu hệ số tương quan $r = r(X, Y)$ giữa hai biến ngẫu nhiên X và Y bằng ± 1, thì $Y = aX + b$ với a, b là các hằng số. Trong trường hợp chung (đặc biệt là khi r^2 gần bằng 1), ta cũng có thể viết Y dưới dạng một đa thức bậc 1 của X cộng với một sai số ϵ nào đó:

$$Y = aX + b + \epsilon. \tag{3.49}$$

Ta muốn chọn các hằng số a và b sao cho sai số ϵ là nhỏ nhất có thể. Ta sẽ dùng chuẩn L_2 để đo độ to nhỏ của ϵ. Có nghĩa là, ta muốn chọn các hằng số a và b sao cho $\mathbb{E}(|\epsilon|^2)$ nhỏ nhất.

Định lý 3.18. *Giả sử X và Y là hai biến ngẫu nhiên không phải hằng số. Đặt $\epsilon = Y - aX - b$ trong đó a, b là hai hằng số thực. Khi đó $\mathbb{E}(|\epsilon|^2)$ đạt giá trị nhỏ nhất (theo a, b) khi mà $a = \dfrac{cov(X,Y)}{var(X)}$ và $b = \mathbb{E}(Y) - a\mathbb{E}(X)$.*

Chứng minh. $\mathbb{E}((Y - aX - b)^2)$ là một đa thức bậc 2 theo a và b, tiến tới $+\infty$ khi $|a| + |b|$ tiến tới vô cùng. Bởi vậy nó đạt giá trị nhỏ nhất tại một điểm mà đạo hàm riêng theo cả hai biến a và b bằng 0. Từ đó ta có hệ phương trình tuyến tính theo a và b:

$$\begin{cases} \frac{\partial \mathbb{E}((Y-aX-b)^2)}{\partial a} = 2a\mathbb{E}(X^2) - 2b\mathbb{E}(X) - 2\mathbb{E}(XY) = 0 \\ \frac{\partial \mathbb{E}((Y-aX-b)^2)}{\partial b} = 2b - 2a\mathbb{E}(X) - 2\mathbb{E}(Y) = 0 \end{cases}. \tag{3.50}$$

Nghiệm duy nhất của hệ phương trình tuyến tính trên là $a = \dfrac{cov(X,Y)}{var(X)}$ và $b = \mathbb{E}(Y) - a\mathbb{E}(X)$, bởi vậy đấy là điểm cực tiểu của $\mathbb{E}((Y - aX - b)^2)$. Có thể tính ra rằng, giá trị cực tiểu của $\mathbb{E}((Y - aX - b)^2)$ bằng $var(Y).(1 - r(X,Y)^2)$. □

Đường thẳng $y = ax + b$ với các hệ số $a = \frac{cov(X,Y)}{var(X)}$ và $b = \mathbb{E}(Y) - a\mathbb{E}(X)$ được gọi là đường **hồi qui tuyến tính** (linear regression), hay đường quan hệ tuyến tính *khớp nhất* (fittest) cho hai biến ngẫu nhiên X và Y, với *sai số bình phương nhỏ nhất*. Đường này là một trường hợp riêng (trường hợp tuyến tính đơn biến) của phương pháp *hồi qui* (regression, mục đích là để viết được một biến ngẫu nhiên dưới dạng hàm số của các biến ngẫu nhiên khác, với sai số chấp nhận được), theo nguyên tắc bình phương nhỏ nhất.

Trong thực tế, ta không biết hết các giá trị của (X, Y) (tức là không biết chính xác phân bố đồng thời của (X, Y)), mà chỉ biết

một số giá trị $(X_1, Y_1), \ldots, (X_n, Y_n)$ của nó (gọi là các **giá trị thực nghiệm**). Khi đó, thay vì không gian xác suất ban đầu của (X, Y), ta có thể sử dụng mô hình **không gian xác suất thực nghiệm** gồm n phần tử, với phân bố xác suất đều, và mỗi phần tử ứng với một cặp giá trị (X_i, Y_i). Ta có thể coi (X, Y) như là vector ngẫu nhiên trên không gian xác suất thực nghiệm này. Khi đó phân bố của (X, Y) trên $\mathbb{R}^2$ cảm sinh bởi không gian xác suất thực nghiệm này được gọi là **phân bố đồng thời thực nghiệm**, và các phân bố biên cảm sinh cũng được gọi là các **phân bố thực nghiệm** (với cỡ của mẫu thực nghiệm bằng n). Việc tính toán hồi qui trong thực tế là dựa trên các phân bố thực nghiệm.

Ví dụ 3.8. Tiếp tục ví dụ 3.6 về quan hệ giữa diện tích căn hộ và giá căn hộ. Có thể tính được rằng (chẳng hạn có thể dùng chương trình gretl để tính), trong ví dụ này, đường quan hệ tuyến tính khớp nhất là đường thẳng $y = 6,14x + 13,7$. Xem hình 3.1. Các điểm (x, y) trong đồ thị phát tán nằm ở hai bên của đường thẳng, và nói chung ở gần đường thẳng. Chú ý rằng, nếu phần lớn các điểm của đồ thị phát tán nằm càng gần đường hồi qui tuyến tính, thì sai số bình phương $\mathbb{E}(|\epsilon|^2)$ càng nhỏ, và hệ số tương quan bình phương r^2 càng gần 1.

Bài tập 3.16. (*Số vụ án mạng, tự sát, và tỷ lệ dân có súng*). Bảng thống kê sau là về số vụ án mạng và số vụ tự sát tính trên 1 triệu dân trong 1 năm, và tỷ lệ số gia đình có súng, ở một số nước trên thế giới, trong các năm 1983-1986, theo số liệu của WHO[3].

[3]Nguồn: M. Killas, International correlation between gun ownership and rates of homicide and suicide, Can. Med. Assoc. J. 1993, 148 (10), 1721–1725.

Nước	Án mạng	Tự sát	% gia đình có súng
Australia	19,5	115,8	19,6
Belgium	18,5	231,5	18,6
Canada	26,0	139,4	29,1
England & Wales	6,7	86,1	4,7
Finland	29,6	253,5	23,2
France	12,5	223, 0	22,6
The Netherlands	11,8	117,2	1,9
Northern Ireland	46,6	82,7	8,4
Norway	12,1	142,7	32,0
Scotland	16,3	105,1	4,7
Spain	13,7	64,5	13,1
Switzerland	11,7	244,5	27,2
United States	75,9	124,0	48,0
West Germany	12,1	203,7	8,9

Dựa vào bảng trên, hãy tính các hệ số tương quan giữa các cặp biến trong 3 biến ngẫu nhiên: "tỷ lệ gia đình có súng", "số vụ án mạng" và "số vụ tự sát", và tính các đường hồi qui tuyến tính giữa của các cặp biến ngẫu nhiên, theo nguyên tắc sai số bình phương nhỏ nhất.

3.4.4 Hệ số tương quan và quan hệ nhân quả

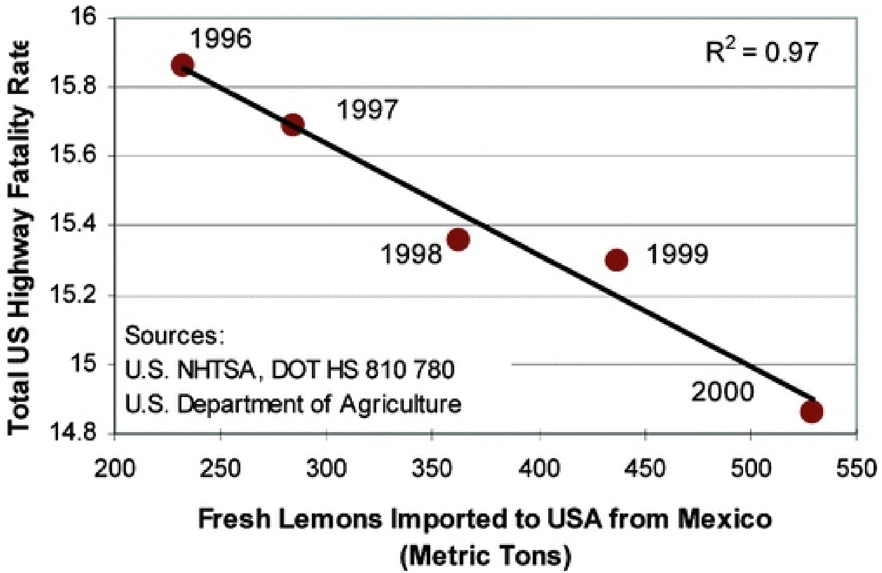
Các biến ngẫu nhiên mà có hệ số tương quan lớn về giá trị tuyệt đối, thường có quan hệ nhân quả (causation) với nhau, liên hệ mật thiết với nhau về logic. Ví dụ, học nhiều thì trình độ cao, trình độ cao

thì dễ xin được việc đòi hỏi trình độ cao. Những việc đòi hỏi trình độ cao, ít người làm được, thì phải trả lương cao để tuyển được người. Từ đó suy ra học nhiều thì thu nhập dễ cao hơn là không có học. Tức là có quan hệ nhân quả giữa "số năm đi học" và "mức thu nhập". Hoặc là ví dụ phía trên về diện tích căn hộ và giá căn hộ: diện tích càng rộng thì ở càng sướng và giá thành cũng càng cao, bởi vậy giá cũng càng cao, tuy rằng tất nhiên có những chỗ diện tích nhỏ lại đắt hơn chỗ khác diện tích rộng hơn, vì giá căn hộ còn phụ thuộc vào những yếu tố khác ngoài diện tích, như là địa điểm, chất lượng nhà, v.v.

Tuy nhiên, không phải lúc nào quan hệ nhân quả cũng rõ ràng. Ví dụ, một nghiên cứu năm 2009 ở đại học Mainz cho thấy ở Đức, "béo phì" (obesity) và "mắc nợ đầm đìa"(over-indebtedness) có tương quan mạnh với nhau[(4)], nhưng không rõ là cái nào dẫn đến cái nào và như thế nào: mắc nợ đầm đìa dẫn đến bị béo phì (do ảnh hưởng tâm lý), hay là bị béo phì dấn đến mắc nợ (do dễ bị mất việc hơn khi béo phì), hay là có những lý do chính khác. Hơn nữa, có những biến ngẫu nhiên mà về mặt logic có thể coi là không liên quan tới nhau, nhưng các giá trị của chúng có hệ số tương quan lớn, do tình cờ. Không gian xác suất càng nhỏ (càng ít phần tử) thì càng dễ xảy ra hiện tượng có các sự kiện không liên quan gì đến nhau nhưng có hệ số tương quan lớn.

Ví dụ 3.9. (Lấy từ Wikipedia[(5)]). Hình 3.3 cho thấy có hệ số tương quan gần bằng -1 giữa số vụ tử vong vì tai nạn xe cộ ở Mỹ trong

[(4)]Xem: http://www.sciencedaily.com/releases/2009/08/090811080751.htm
[(5)]Xem trang web http://en.wikipedia.org/wiki/Correlation_does_not_imply_causation



Hình 3.3: Tai nạn giao thông chết người và chanh tươi nhập khẩu

những năm 1996-2000, và số quả chanh nhập khẩu sang Mỹ từ Mexico. Tuy nhiên người ta có thể tự hỏi: hai biến đó thì liên quan gì với nhau?!

3.5 Phân bố và kỳ vọng có điều kiện

Ở phần này chúng ta sẽ bàn về phân bố xác suất và kỳ vọng của một biến ngẫu nhiên X dưới điệu kiện $Y = y$, trong đó y là một số và Y là một biến ngẫu nhiên khác. Trong trường hợp Y có phân bố xác suất liên tục tại y, tức là $P(Y = y) = 0$, thì chúng ta không thể định nghĩa hàm phân phối xác suất có điều kiện $P(X \leq x|Y = y)$

theo công thức xác suất có điều kiện thông thường, $P(X \leq x|Y = y) = P((X \leq x) \cap (Y = y))/P(Y = y)$, mà chúng ta sẽ phải dùng các phương pháp giới hạn giải tích để định nghĩa và nghiên cứu nó. Còn trong trường hợp biến ngẫu nhiên Y có $P(Y = y) > 0$, thì ta có thể dùng công thức xác suất có điều kiện thông thường.

3.5.1 Trường hợp rời rạc

Định nghĩa 3.8. *Giả sử X, Y là hai biến ngẫu nhiên, $y \in \mathbb{R}$, và $P(Y = y) > 0$. Khi đó* **phân bố xác suất có điều kiện** *của X với điều kiện $Y = y$ là phân bố xác suất trên $\mathbb{R}$ cho bởi công thức sau:*

$$P_{X|Y=y}(A) = P_{X|Y}(A|y) = P(X \in A|Y = y) = \frac{P(X \in A, Y = y)}{P(Y = y)} \tag{3.51}$$

(với mọi tập hợp $A \subset \mathbb{R}$ thuộc sigma-đại số Borel). **Hàm phân phối xác suất có điều kiện** *là hàm*

$$\mathcal{F}_{X|Y=y}(x) = \mathcal{F}_{X|Y}(x|y) = P_{X|Y=y}(]-\infty, x]) = \frac{P(X \leq x, Y = y)}{P(Y = y)}. \tag{3.52}$$

Kỳ vọng có điều kiện *của X với điều kiện $Y = y$ là*

$$\mathbb{E}(X|Y = y) = \int_{x \in \mathbb{R}} x dP_{X|Y=y}. \tag{3.53}$$

Nói cách khác, kỳ vọng có điều kiện chính là kỳ vọng của phân bố xác suất $P_{X|Y=y}$ trên $\mathbb{R}$. Trong trường hợp phân bố xác suất $P_{X|Y=y}$ là rời rạc và tập trung tại các điểm $x_1, x_2, \ldots$, thì kỳ vọng có điều kiện

có thể được tính theo công thức

$$\mathbb{E}(X|Y=y) = \sum_i x_i.P_{X|Y=y}(x_i) = \sum_i x_i.P(X=x_i|Y=y), \tag{3.54}$$

với $P(X=x_i|Y=y) = \dfrac{P(X=x_i, Y=y)}{P(Y=y)} = \dfrac{P_{X,Y}(x_i,y)}{P_Y(y)}$.

Ví dụ 3.10. Giả sử X và Y là hai biến ngẫu nhiên độc lập với phân bố Poisson với các tham số λ và γ tương ứng. Chúng ta sẽ tính phân bố xác suất có điều kiện của X với điều kiện $X+Y$ cho trước, tức là tính $P(X=k|X+Y=k+m)$. Tổng $X+Y$ cũng có phân bố Poisson với tham số $\lambda+\gamma$. Bởi vậy,

$$\begin{aligned}
P(X=k|X+Y=k+m) &= \frac{P(X=k, Y=m)}{P(X+Y=k+m)} \\
&= \frac{\frac{\lambda^k}{k!}e^{-\lambda}\frac{\gamma^m}{m!}e^{-\gamma}}{\frac{(\lambda+\gamma)^{k+m}}{(k+m)!}e^{-(\lambda+\gamma)}} = \frac{(k+m)!}{k!m!}\frac{\lambda^k\gamma^m}{(\lambda+\gamma)^{k+m}} \\
&= C_{k+m}^k(\frac{\lambda}{\lambda+\gamma})^k(1-\frac{\lambda}{\lambda+\gamma})^m
\end{aligned}$$

Nói cách khác, đặt $r=k+m$, ta có

$$P_{X|X+Y=r}(k) = C_r^k(\frac{\lambda}{\lambda+\gamma})^k(1-\frac{\lambda}{\lambda+\gamma})^{r-k},$$

có nghĩa là phân bố xác suất $P_{X|X+y=r}$ là phân bố nhị thức với các tham số r và $\frac{\lambda}{\lambda+\gamma}$. Từ đó suy ra $\mathbb{E}(X|X+Y=r) = \dfrac{r\lambda}{\lambda+\gamma}$.

Kỳ vọng có điều kiện có thể được dùng để tính kỳ vọng không điều kiện qua công thức sau:

Định lý 3.19. *Giả sử Y là một biến ngẫu nhiên rời rạc và X là một*

biến ngẫu nhiên. Khi đó

$$\mathbb{E}(X) = \sum_y \mathbb{E}(X|Y = y)P(Y = y). \qquad (3.55)$$

Chứng minh. Ta có thể viết $X = \sum_y X_y$, với $X_y = X$ khi mà $Y = y$ và $X_y = 0$ khi mà $Y \neq y$. Khi đó $\mathbb{E}(X|Y = y) = \mathbb{E}(X_y)/P(Y = y)$, và

$$\mathbb{E}(X) = \mathbb{E}(\sum_y X_y) = \sum_y \mathbb{E}(X_y) = \sum_y \mathbb{E}(X|Y = y)P(Y = y).$$

□

Ví dụ 3.11. Giả sử một cửa hàng bán một loại đồ chơi đặc biệt. Mỗi khách hàng trong ngày có xác suất mua đồ chơi đặc biệt là p, và các quyết định mua của các khách hàng là độc lập với nhau. Số khách hàng trong ngày là một số ngẫu nhiên N tuân theo phân bố Poisson với tham số λ. Gọi K là số khách hàng mua đồ chơi đặc biệt trong ngày. Chúng ta muốn tính $\mathbb{E}(K|N = n)$ và $\mathbb{E}(K)$. Các giả sử phía trên cho biết $P(N = n) = \lambda^n e^{-\lambda}/n!$ (phân bố Poisson), và $P_{K|N}(k|n) = C_n^k p^k (1-p)^k$ (phân bố nhị thức). Từ đó suy ra $\mathbb{E}(K|N = n) = pn$, và $\mathbb{E}(K) = \sum_n \mathbb{E}(K|N = n)P(N = n) = p\sum_n nP(N = n) = p\mathbb{E}(N) = p\lambda$.

Bài tập 3.17. Chứng minh rằng, trong ví dụ 3.11, biến K tuân theo phân bố Poisson với tham số $p\lambda$, và $\mathbb{E}(N|K = k) = k + \lambda(1-p)$.

Bài tập 3.18. Giả sử X có phân bố mũ với tham số $\lambda > 0$. Giả sử Y là biến ngẫu nhiên sao cho khi $X = x$ thì Y có phân bố đều trên đoạn thẳng $[0, x]$. Hãy tính $\mathbb{E}(Y)$.

3.5.2 Trường hợp liên tục

Khi $P(Y = y) = 0$, ta không định nghĩa được $P(X \leq x|Y = y)$ một cách trực tiếp như trong trường hợp $P(Y = y) > 0$, mà phải dùng đến các phép toán giải tích có sử dụng giới hạn. Một trong các định nghĩa có thể dùng là:

$$\mathcal{F}_{X|Y=y}(x) := P(X \leq x|Y = y) := \lim_{\epsilon\to 0+} P(X \leq x|y \leq Y \leq y + \epsilon), \tag{3.56}$$

nếu như giới hạn trên tồn tại. Trong trường hợp rời rạc, có thể chứng minh rằng giới hạn trên luôn tồn tại và cho kết quả trùng với định nghĩa thông thường. Ở đây chúng ta sẽ chỉ quan tâm đến những trường hợp liên tục "đủ tốt" sao cho giới hạn trên tồn tại.

Giả sử vector ngẫu nhiên (X, Y) có hàm mật độ đồng thời $\rho_{X,Y}$ và các hàm mật độ biên ρ_X, ρ_Y. Khi đó ta có thể viết:

$$P(X \leq x|Y = y) = \lim_{\epsilon\to 0+} P(X \leq x|y \leq Y \leq y + \epsilon) =$$

$$= \lim_{\epsilon\to 0+} \frac{P(X \leq x, y \leq Y \leq y + \epsilon)}{P(\leq Y \leq y + \epsilon)} = \lim_{\epsilon\to 0+} \frac{\int_{-\infty}^{x}(\int_{y}^{y+\epsilon} \rho_{X,Y}(t, s)ds)dt}{\int_{x}^{x+\epsilon} \rho_Y(s)ds} =$$

$$= \frac{\int_{-\infty}^{x} \lim_{\epsilon\to 0+}(\int_{y}^{y+\epsilon} \rho_{X,Y}(t, s)ds/\epsilon)dt}{\lim_{\epsilon\to 0+}(\int_{x}^{x+\epsilon} \rho_Y(s)ds/\epsilon)} = \int_{-\infty}^{x} \frac{\rho_{X,Y}(t, y)}{\rho_Y(y)}dt,$$

và bởi vậy, ta có:

Mệnh đề 3.20. *Trong trường hợp liên tục tuyệt đối, nếu $\rho_Y(y) > 0$ thì hàm mật độ của phân bố xác suất có điều kiện $P_{X|Y=y}$ chính là hàm*

$$\rho_{X|Y}(x|y) := \rho_{X|Y=y}(x) = \frac{\rho_{X,Y}(x, y)}{\rho_Y(y)}. \tag{3.57}$$

Kỳ vọng có điều kiện trong trường hợp liên tục có thể được viết dưới dạng:

$$\mathbb{E}(X|Y=y) = \int_{\mathbb{R}} x dP_{X|Y=y} = \int_{-\infty}^{\infty} x\rho_{X|Y}(x|y)dx. \qquad (3.58)$$

Định lý 3.21. *Ta có các công thức sau:*

i)

$$\mathcal{F}_X(x) = P(X \le x) = \int_{-\infty}^{\infty} \mathcal{F}_{X|Y=y}(x)\rho_Y(y)dy, \qquad (3.59)$$

ii)

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \mathbb{E}(X|Y=y)\rho_Y(y)dy. \qquad (3.60)$$

Chứng minh. Kiểm tra công thức thứ hai:

$$\begin{aligned}\mathbb{E}(X) = \int_{y=-\infty}^{\infty}\int_{x=-\infty}^{\infty} x\rho_{X,Y}(x,y)dxdy \\ = \int_{y=-\infty}^{\infty}\left(\int_{x=-\infty}^{\infty} x\rho_{X|Y}(x|y)\rho_Y(y)dx\right)dy \\ = \int_{-\infty}^{\infty}\mathbb{E}(X|Y=y)\rho_Y(y)dy.\end{aligned}$$

Chứng minh của công thức thứ nhất hoàn toàn tương tự, và nhường cho bạn đọc làm bài tập. Công thức thứ nhất cũng có thể được suy từ công thức thứ hai, bằng cách thay biến ngẫu nhiên X bằng biến ngẫu nhiên $\mathcal{X}_x$ định nghĩa bởi: $\mathcal{X}_x = 1$ khi $X \le x$ và $\mathcal{X}_x = 0$ khi $X > x$. (Khi đó $\mathbb{E}(\mathcal{X}_x) = \mathcal{F}_X(x)$). □

Ví dụ 3.12. Xét vector ngẫu nhiên liên tục (X, Y) với hàm mật độ: $\rho_{X,Y}(x,y) = 1/x$ khi $0 < y \le x \le 1$, và $\rho_{X,Y}(x,y) = 0$ tại các điểm khác. Dễ thấy rằng, với mỗi x cố định, $0 < x \le 1$, phân bố xác suất có điều kiện $P_{Y|X=x}$ là phân bố đều trên đoạn thẳng $]0, x]$, với hàm mật

độ bằng $1/x$ trên đoạn thẳng đó. Phân bố xác suất biên P_X là phân bố đều trên đoạn $]0,1]$. Từ đó suy ra $\mathbb{E}(Y) = \int_0^1 \mathbb{E}(Y|X=x)\rho_X(x)dx = \int_0^1 \mathbb{E}(Y|X=x)dx = \int_0^1 (x/2)dx = 1/4$.

3.6 Phân bố normal nhiều chiều

3.6.1 Định nghĩa của phân bố normal nhiều chiều

Phân bố normal n chiều ($n \geq 2$) là mở rộng của khái niệm phân bố normal trên $\mathbb{R}$ lên trường hợp nhiều chiều, và đóng vai trò rất quan trọng trong việc nghiên cứu các quá trình ngẫu nhiên (mà trong khuôn khổ của quyển sách này chúng ta không bàn tới). Ví dụ đơn giản nhất của phân bố normal nhiều chiều là, nếu $Z_1, \ldots, Z_n$ là một bộ n biến ngẫu nhiên độc lập với nhau và cùng tuân theo phân bố normal chuẩn tắc $\mathcal{N}(0,1)$, thì phân bố đồng thời của $(Z_1, \ldots, Z_n)$, với hàm mật độ đồng thời

$$\rho(x_1, \ldots, x_n) = \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\frac{\sum_i x_i^2}{2}\right), \tag{3.61}$$

là một phân bố normal nhiều chiều, gọi là **phân bố normal nhiều chiều chuẩn tắc**. Tương tự như trong trường hợp một chiều, ta muốn rằng một biến đổi tuyến tính (hay affine, tức là tuyến tính cộng với một phép tịnh tiến) của một phân bố normal nhiều chiều cũng là một phân bố normal nhiều nhiều. Bởi vậy ta có định nghĩa sau:

Định nghĩa 3.9. *Ta nói rằng một vector ngẫu nhiên $\mathbf{X} = (X_1, \ldots, X_n)$ có* **phân bố normal n chiều**, *nếu như tồn tại một bộ m biến ngẫu nhiên $\mathbf{Z} = (Z_1, \ldots, Z_m)$ độc lập ($m \in \mathbb{N}$), với các Z_i cùng tuân theo*

phân bố normal chuẩn tắc $\mathcal{N}(0,1)$, một ma trận $A = (a_{ij})_{i=1,\ldots,n}^{j=1,\ldots,m}$ và một vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ (A và $\boldsymbol{\mu}$ là hằng số), sao cho:

$$\mathbf{X}^t = A.\mathbf{Z}^t + \boldsymbol{\mu}^t. \quad (3.62)$$

Chữ t ở trong công thức trên là phép chuyển vị ma trận, để biến các vector hàng thành vector cột. Nói cách khác,

$$X_i = \sum_{k=1}^{m} a_{ik} Z_k + \mu_i \quad \forall i = 1, \ldots, n. \quad (3.63)$$

Tương tự như trong trường hợp 1 chiều, các phân bố normal nhiều chiều có thể dùng làm mô hình xác suất của khá nhiều vấn đề thực tế. Ví dụ, bộ 3 biến (chiều cao của một người đàn bà, cân nặng của người đó, chỉ số trí tuệ của người đó) có thể được coi là 1 vector ngẫu nhiên 3 chiều với phân bố normal 3 chiều. Cơ sở toán học để giải thích điều này cũng là định lý giới hạn trung tâm.

Vì tổng của các biến ngẫu nhiên độc lập với phân bố normal cũng là biến ngẫu nhiên với phân bố normal, nên nếu $\mathbf{X} = (X_1, \ldots, X_n)$ có phân bố normal n chiều, thì các thành phần X_i của nó đều có phân bố normal, tuy điều ngược lại không đúng.

Ma trận đối xứng $\Sigma = A.A^t$, trong đó A là ma trận trong định nghĩa trên, được gọi là **ma trận hiệp phương sai** của phân bố normal nhiều chiều trong định nghĩa. Lý do là vì phần tử Σ_{ij} của ma trận Σ chính là hiệp phương sai $cov(X_i, X_j)$ của X_i và X_j. Thật vậy, theo định nghĩa, ta có $X_i = \sum_k a_{ik} Z_k + \mu_i$, với kỳ vọng bằng μ_i. Từ đó suy ra $cov(X_i, X_j) = \mathbb{E}((\sum_k a_{ik} Z_k)(\sum_k a_{jk} Z_k)) = \sum_k a_{ik} a_{jk} = \Sigma_{ij}$.

Vector $\boldsymbol{\mu}$ được gọi là **vector kỳ vọng** của phân bố normal nhiều chiều trên. Một phân bố normal nhiều chiều được xác định duy nhất

bởi ma trận hiệp phương sai Σ và vector kỳ vọng $\boldsymbol{\mu}$ của nó, và thường được ký hiệu là $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$.

3.6.2 Trường hợp hai chiều

Để hiểu hơn về phân bố normal nhiều chiều, trước hết chúng ta sẽ xét kỹ hơn trường hợp 2 chiều. Gọi (X_1, X_2) là một vector ngẫu nhiên 2 chiều với phân bố normal. Theo định nghĩa, ta có:

$$X_1 = \sum_{i=1}^{m} a_{1i} Z_i + \mu_1, \quad X_2 = \sum_{i=1}^{m} a_{2i} Z_i + \mu_1, \tag{3.64}$$

trong đó $Z_1, \ldots, Z_m$ là một bộ m biến ngẫu nhiên độ lập có phân bố normal chuẩn tắc $\mathcal{N}(0, 1)$.

Ma trận hiệp phương sai Σ trong trường hợp này là ma trận 2×2, với 4 phần tử:

$$\Sigma_{11} = \sum_k a_{1k}^2, \Sigma_{12} = \Sigma_{21} = \sum_k a_{1k} a_{2k}, \Sigma_{22} = \sum_k a_{2k}^2. \tag{3.65}$$

Bổ đề 3.22. *Nếu ma trận hiệp phương sai Σ là ma trận đường chéo, tức là $\Sigma_{12} = \Sigma_{21} = cov(X_1, X_2) = 0$ thì hai biến ngẫu nhiên X_1 và X_2 độc lập với nhau.*

Chứng minh. Có thể kiểm tra trực tiếp theo định nghĩa, hoặc là dùng hàm đặc trưng, kiểm tra rằng $\Phi_{(X_1,X_2)}(s_1, s_2) = \Phi_{X_1}(s_1).\Phi_{X_1}(s_2)$ nếu như $cov(X_1, X_2) = 0$. Chú ý rằng, nếu X_1, X_2 là hai biến ngẫu nhiên tùy ý, thì từ $cov(X_1, X_2) = 0$ không suy ra được rằng X_1 độc lập với X_2. Nhưng ở đây X_1 và X_2 là hai thành phần của một vector ngẫu nhiên với phân bố normal, nên điều đó đúng. □

Nếu Σ không có dạng đường chéo ($\Sigma_{12} \neq 0$), thì vì Σ đối xứng nên ta có thể đường chéo hóa nó bằng một ma trận 2×2 vuông góc (orthogonal) C (vuông góc có nghĩa là $C^{-1} = C^t$): $\Sigma' = C^{-1}\Sigma C = C^t \Sigma C$ là ma trận đường chéo. Đặt $(Y_1, Y_2)^t = C^{-1}.((X_1, X_2)^t - \boldsymbol{\mu}^t) = (C^{-1}A).\mathbf{Z}^t$. Khi đó (Y_1, Y_2) có phân bố normal, và có ma trận hiệp phương sai bằng $(C^{-1}A)(C^{-1}A)^t = C^{-1}\Sigma C = \Sigma'$, là một mà trận đường chéo, và bởi vậy Y_1 và Y_2 độc lập với nhau. Ta có thể viết $Y_1 = \beta_1 Z'_1$, $Y_2 = \beta_2 Z'_2$, trong đó α_1 và α_2 là độ lệch chuẩn của Y_1 và Y_2, và Z'_1 và Z'_2 độc lập và có phân bố normal chuẩn tắc $\mathcal{N}(0, 1)$. Vì $(X_1, X_2)^t = C.(Y_1, Y_2)^t + \boldsymbol{\mu}^t)$ nên ta có thể viết:

$$\begin{cases} X_1 = a'_{11}.Z'_1 + a'_{12}.Z'_2 + \mu_1 \\ X_2 = a'_{21}.Z'_1 + a'_{22}.Z'_2 + \mu_2 \end{cases}. \tag{3.66}$$

Có nghĩa là, trong trường hợp vector 2 chiều, ta luôn có thể giả sử $m = 2$: Để sinh ra một vector ngẫu nhiên 2 chiều với phân bố normal tùy ý, chỉ cần biến đổi tuyến tính một vector ngẫu nhiên 2 chiều với phân bố normal chuẩn tắc.

Trường hợp $\det \Sigma = 0$ gọi là **trường hợp suy biến**. Khi đó (ít nhất) một trong hai giá trị riêng (eigenvalue) của Σ bằng 0, suy ra một trong hai biến ngẫu nhiên, Y_1 và Y_2 phía trên là hằng số, và khi đó thực ra ta chỉ cần một biến ngẫu nhiên với phân bố normal chuẩn tắc $\mathcal{N}(0, 1)$ để sinh ra vector (X_1, X_2). Nói cách khác, trong trường hợp suy biến, ta có thể viết

$$\begin{cases} X_1 = \alpha_1 Z + \mu_1 \\ X_2 = \alpha_2 Z + \mu_2 \end{cases} \tag{3.67}$$

với α_1, α_2 là hằng số, và Z là một biến ngẫu nhiên có phân bố normal

chuẩn tắc $\mathcal{N}(0,1)$. Phân bố đồng thời của (X_1, X_2) trong trường hợp suy biến tập trung trên đường thẳng $\alpha_2(x_1 - \mu_1) = \alpha_1(x_2 - \mu_2)$ trong $\mathbb{R}^2$, và bởi vậy nó không có hàm mật độ trên $\mathbb{R}^2$.

Trường hợp $\det \Sigma \neq 0$ gọi là **trường hợp không suy biến**. Khi đó phân bố xác suất của $\mathbf{X} = (X_1, X_2)$ có hàm mật độ sau:

$$\rho_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}).\Sigma^{-1}.(\mathbf{x} - \boldsymbol{\mu})^t). \tag{3.68}$$

Thật vậy, nếu thay vì xét hệ tọa độ $\mathbf{x} = (x_1, x_2)$ trên $\mathbb{R}^2$, ta xét hệ tọa độ mới $\mathbf{y} = (y_1, y_2)$, qua phép biến đổi affine $\mathbf{y}^t = C^{-1}.(\mathbf{x}^t - \boldsymbol{\mu}^t)$, thì ta có

$$\rho_{\mathbf{X}}(\mathbf{x}) = \rho_{\mathbf{Y}}(\mathbf{y}) \tag{3.69}$$

(hàm mật độ không thay đổi, vì phép đổi biến bảo toàn diện tích Euclid), và

$$\begin{aligned}
\rho_{\mathbf{Y}}(\mathbf{y}) = \rho_{Y_1}(y_1)\rho_{Y_2}(y_1) &= \frac{1}{2\pi\beta_1\beta_2} \exp(-\frac{1}{2}(\frac{y_1^2}{\beta_1^2} + \frac{y_2^2}{\beta_2^2})) \\
&= \frac{1}{2\pi\sqrt{\det \Sigma'}} \exp(-\frac{1}{2}(\mathbf{y}.(\Sigma')^{-1}.\mathbf{y}^t) \\
&= \frac{1}{2\pi\sqrt{\det \Sigma}} \exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}).\Sigma^{-1}.(\mathbf{x} - \boldsymbol{\mu})^t).
\end{aligned}$$

Các đường mức của hàm mật độ $\rho_{\mathbf{X}}(\mathbf{x})$ của phân bố normal hai chiều trên $\mathbb{R}^2$ là các đường ellipse, với tâm điểm tại $\mathbf{x} = \mathbf{m}$ (không điểm của hệ tọa độ (y_1, y_2)) và các trục là các trục của hệ tọa độ (y_1, y_2).

Ví dụ 3.13. Hàm

$$\rho_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1 - r^2}} \exp\left(-\frac{1}{2(1 - r)^2}(x^2 - 2rxy + y^2)\right), \tag{3.70}$$

là hàm mật độ của một phân bố normal hai chiều (bivariate normal distribution) với tham số r, $-1 < r < 1$. Tham số r ở đây chính là hệ số tương quan giữa hai thành phần X và Y. Các phân bố biên P_X và P_Y của phân bố normal hai chiều này là các phân bố normal chuẩn tắc $\mathcal{N}(0,1)$. Ma trận hiệp phương sai ở đây là mà trận

$$\Sigma = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}. \tag{3.71}$$

Hai thành phần X và Y ở đây độc lập với nhau khi và chỉ khi $r = 0$.

3.6.3 Một số tính chất của phân bố normal nhiều chiều

Định lý 3.23. *i) Giả sử một vector ngẫu nhiên n chiều $\mathbf{X}$ có phân bố normal. Khi đó phân bố của nó được xác định duy nhất bởi ma trận hiệp phương sai Σ và vector kỳ vọng $\boldsymbol{\mu}$ của nó. Nói cách khác, hai phân bố normal n chiều có cùng ma trận hiệp phương sai và vector kỳ vọng thì bằng nhau.*
ii) Nếu hạng của ma trận hiệp phương sai Σ bằng k ($k \leq n$), thì $\mathbf{X}$ có thể được sinh bởi một họ k biến ngẫu nhiên độc lập có phân bố chuẩn $\mathcal{N}(0,1)$ qua một phép biến đổi affine, và phân bố của $\mathbf{X}$ tập trung tại một không gian affine con có số chiều bằng k trên $\mathbb{R}^n$.
iii) Nếu ma trận hiệp phương sai Σ là không suy biến (tức là $\det \Sigma \neq 0$, hay nói cách khác, hạng của Σ bằng n), thì phân bố normal $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ có hàm mật độ $\rho_{\mathbf{X}}$ sau trên $\mathbb{R}^n$ (ở đây ta sử dụng các ký hiệu $\mathbf{X} = (X_1, \ldots, X_n)$ $\mathbf{x} = (x_1, \ldots, x_n)$, và $|\Sigma| = \det \Sigma$) :

$$\rho_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}).\Sigma^{-1}.(\mathbf{x} - \boldsymbol{\mu})^t). \tag{3.72}$$

Các mặt mức của hàm mật độ $\rho_{\mathbf{X}}$ *là các hình ellipsoid đồng dạng có tâm điểm tại* $\boldsymbol{\mu}$*. Nếu ma trận* Σ *suy biến thì phân bố* $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ *không có hàm mật độ.*
iv) Với mọi $\mathbf{c} = (c_1, \ldots, c_n) \in \mathbb{R}^n$*, biến ngẫu nhiên* $X = \sum_{i=1}^n c_i X_i$ *có phân bố normal* $\mathcal{N}(\mu, \sigma^2)$*, với* $\mu = \mathbb{E}(X) = \sum_{i=1}^n c_i \mu_i$*, và* $\sigma^2 = var(\sum_{i=1}^n c_i X_i) = \mathbf{c}.\Sigma.\mathbf{c}^t$*. (Nếu* $var(\sum_{i=1}^n c_i X_i) = 0$ *thì* X *là hằng số, có nghĩa là phân bố xác suất của* X *tập trung tại một điểm).*
v) Ngược lại, giả sử rằng $(X_1, \ldots, X_n)$ *là một vector ngẫu nhiên với tính chất: phân bố xác suất của* $\sum_{i=1}^n c_i X_i$ *là phân bố normal (hoặc là tập trung tại một điểm) với mọi* $(c_1, \ldots, c_n) \in \mathbb{R}^n$*. Khi đó, phân bố xác suất đồng thời của* $(X_1, \ldots, X_n)$ *là một phân bố normal* n *chiều.*

Trong mục trước, chúng ta đã chứng minh về cơ bản định lý trên trong trường hợp 2 chiều, trừ khẳng định cuối cùng. Trường hợp tổng quát n chiều chứng minh hoàn toàn tương tự trường hợp 2 chiều. Khẳng định cuối cùng có thể chứng minh bằng cách xét hàm đặc trưng. □

Bài tập 3.19. Chứng minh rằng phân bố normal n chiều $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ có hàm đặc trưng sau:

$$\Phi(\mathbf{s}) = \exp(\sqrt{-1}\boldsymbol{\mu}.\mathbf{s}^t - \frac{1}{2}\mathbf{s}.\Sigma.\mathbf{s}^t). \tag{3.73}$$

Bài tập 3.20. Giả sử X và Y là hai biến ngẫu nhiên độc lập tuân theo phân bố normal chuẩn tắc $\mathcal{N}(0, 1)$. Tính hàm mật độ của $Z = \frac{1}{2}(X^2 + Y^2)$.

Bài tập 3.21. Ta gọi **phân bố Cauchy** là phân bố liên tục trên $\mathbb{R}$ với hàm mật độ sau:

$$\rho(x) = \frac{1}{\pi(1 + x^2)}. \tag{3.74}$$

(Phân bố này không có kỳ vọng, và không có phương sai hữu hạn). Chứng minh rằng nếu Z_1 và Z_2 là hai biến ngẫu nhiên độc lập tuân theo phân bố normal chuẩn tắc $\mathcal{N}(0,1)$, thì Z_1/Z_2 có phân bố Cauchy.

Bài tập 3.22. Giả sử vector (X,Y) có phân bố normal 2 chiều với hàm mật độ

$$\rho_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-r^2}}\exp(-\frac{1}{2(1-r)^2}(x^2-2rxy+y^2)).$$

i) Chứng minh rằng X và $Z = (Y-rX)/(1-r^2)^{1/2}$ là các biến ngẫu nhiên độc lập có phân bố normal chuẩn tắc $\mathcal{N}(0,1)$.
ii) Suy ra từ i) rằng

$$P(X>0,Y>0) = \frac{1}{4}+\frac{1}{2\pi}\arcsin r.$$

iii) Chứng minh rằng với mọi $y \in \mathbb{R}$, phân bố xác suất có điều kiện $P_{X|Y=y}$ là một phân bố normal có phương sai không phụ thuộc vào điểm y, và tính phương sai và kỳ vọng của phân bố có điều kiện đó.

Chương 4

Các định lý giới hạn

4.1 Định lý giới hạn trung tâm

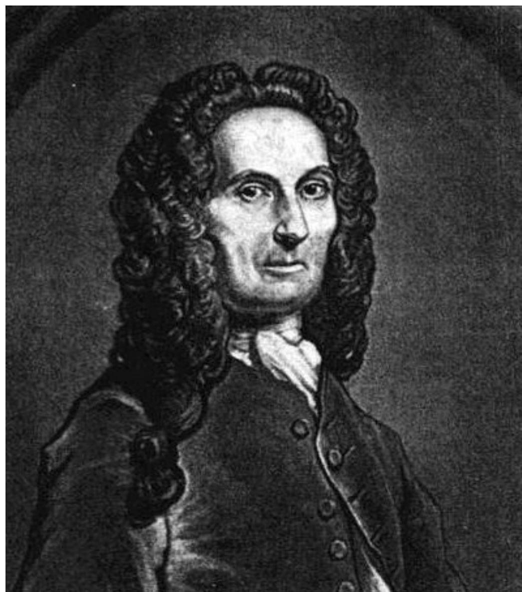
Định lý giới hạn trung tâm là được coi là định lý quan trọng nhất của xác suất thống kê, *hòn đá tảng* của thống kê toán học. Nó là một trong những định lý được trích dẫn sử dụng nhiều nhất của toàn bộ toán học hiện đại nói chung.

4.1.1 Định lý de Moivre – Laplace

Tiền thân của định giới hạn trung tâm tổng quát là định lý sau đây của de Moivre và Laplace về dáng điệu tiệm cận của phân bố xác suất nhị thức

$$P_n(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \tag{4.1}$$

với tham số p cố định, khi n tiến tới vô cùng.



Hình 4.1: Abraham de Moivre (1667–1754)

Định lý 4.1 (de Moivre – Laplace). *Đặt*

$$z = z(n,k) = (k - np)/\sqrt{np(1-p)}. \tag{4.2}$$

Khi đó

$$P_n(k) = \frac{1}{\sqrt{2\pi}\sqrt{np(1-p)}} \exp(-\frac{1}{2}z^2).(1 + \delta_n(k)), \tag{4.3}$$

trong đó $\delta_n(k)$ hội tụ đều đến 0 khi n tiến tới ∞, có nghĩa là

$$\lim_{n\to\infty} \sup_k \delta_n(k) = 0.$$

Định lý 4.1 liên quan chặt chẽ đến **công thức Sterling** sau đây trong giải tích:

$$\lim_{n\to+\infty} \frac{n!}{\sqrt{2\pi n}\,\left(\frac{n}{e}\right)^n} = 1. \tag{4.4}$$

Sử dụng công thức Sterling, có thể chứng minh khá dễ dàng định lý de Moivre – Laplace 4.1, và ngược lại, công thức Sterling cũng có thể suy được ra từ định lý 4.1. Ở đây chúng ta sẽ tạm thời chấp nhận định lý 4.1 và công thức Sterling mà không chứng minh[1].

Một hệ quả trực tiếp và quan trọng của định lý 4.1 là định lý sau:

Định lý 4.2 (de Moivre – Laplace). *Giả sử* $X_1, X_2, \ldots, X_n, \ldots$ *là các biến ngẫu nhiên độc lập có cùng phân bố xác suất Bernoulli:* $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$ *với mọi* i*. Đặt* $S_n = X_1 + \ldots + X_n$*. Khi đó với mọi cặp số thực* $a < b$ *ta có:*

$$\lim_{n\to\infty} P\left(a \le \frac{S_n - pn}{\sqrt{np(1-p)}} \le b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \tag{4.5}$$

Chứng minh. Theo giả thuyết, S_n có phân bố nhị thức $P(S_n = k) = P_n(k) = C_n^k p^k (n-p)^{n-k}$. Đặt $z = z(n,k) = \dfrac{k - pn}{\sqrt{np(1-p)}}$. Áp

[1] Các chứng minh cổ điển của công thức Sterling khá dài. Nhưng có thể xem một chứng minh ngắn gọn và đơn giản, dựa trên hàm gamma và định lý hội tụ bị chặn Lebesgue (định lý 2.8) trong bài báo sau: J. M. Patin, A very short proof of Sterling's formula, The American Mathematical Monthly, Vol. 96 (1989), No. 1, pp 41–42.

dụng định lý 4.1, ta có:

$$P\left(a \le \frac{S_n - pn}{\sqrt{np(1-p)}} \le b\right) = \sum_{a \le z \le b} P_n(k)$$
$$= \sum_{a \le z \le b} \Delta_n \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2).(1+\delta_n(k)), \quad (4.6)$$

trong đó $\Delta_n = \dfrac{1}{\sqrt{np(1-p)}}$ bằng bước nhảy của z trong tổng phía trên. Bởi vậy, theo định nghĩa tích phân Riemann, ta có

$$\lim_{n\to\infty} \sum_{a \le z \le b} \Delta_n \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx, \quad (4.7)$$

từ đó suy ra điều phải chứng minh. □

Định lý 4.2 chính là một trường hợp riêng quan trọng của định lý giới hạn trung tâm bàn đến ở mục sau.

Ví dụ 4.1. Tung một đồng tiền 1000 lần, có 600 lần hiện mặt ngửa. Ta có thể coi đồng tiền là cân bằng (hai mặt sấp và ngửa đều có xác suất hiện lên là 1/2) được không? Để trả lời câu hỏi đó, ta giả sử là đồng tiền cân bằng. Khi đó ta có phân bố nhị thức với $n = 1000$, $p = 1/2$, $pn = 500$, $\sqrt{np(1-p)} \approx 15,1811$ Gọi k là số lần hiện lên mặt ngửa trong số $n = 400$ lần tung. Theo định lý de Moivre – Laplace, ta có $P(k \le 599) = P\left(\frac{k-np}{\sqrt{np(1-p)}} \le \frac{99}{15,1811}\right) \approx \int_{-\infty}^{6,521} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx >$ $0,9999999999$. Điều đó có nghĩa là, nếu đồng xu cân bằng, thì xác suất để hiện lên mặt ngửa ít nhất 600 lần khi tung đồng xu 1000 lần nhỏ hơn $1/10^{10}$. Khả năng xảy ra điều đó là quá nhỏ để có thể tin được là đồng xu cân bằng.

Ghi chú 4.1. Abraham de Moivre (1667–1754) là một nhà toán học người gốc Pháp, bị bắt đi tù năm 1688 vì lý do tôn giáo, sau đó di tản sang London và ở đó cho đến khi chết. Được bầu vào viện Hàm lâm Hoàng gia Anh (Royal Society) năm 1697. Cùng với Newton và Leibniz, de Moivre là một trong những người đầu tiên nghiên cứu phép tính vi phân (differential calculus), mà thời đó gọi là *method of fluxions*. Khi người ta hỏi Newton về method of fluxions, Newton có khẳng định là "nên gặp de Moivre vì ông ta biết tốt hơn tôi". Định lý de Moivre–Laplace về dáng điệu tiệm cận của phân bố nhị thức đầu tiên là do de Moivre phát hiện và chứng minh cho trường hợp $p = 1/2$ từ năm 1733, sau đó nó được Laplace mở rộng cho trường hợp p bất kỳ. Ngoài lý thuyết xác suất và phép tính vi phân, de Moivre còn là một trong những người đầu tiên nghiên cứu lý thuyết tập hợp và số phức. Công thức $(\cos(x) + i\sin(x))^n = \cos(nx) + i\sin(nx)$ cho số phức mang tên công thức de Moivre.

Bài tập 4.1. Tính xác suất của sự kiện sau: tung một con xúc sắc (đều) 6000 lần, số lần xuất hiện mặt 6 là một số ≥ 850 và ≤ 1050.

4.1.2 Định lý giới hạn trung tâm

Giả sử $X_1, X_2, \ldots, X_n, \ldots$ là một dãy các biến ngẫu nhiên độc lập có cùng phân bố xác suất, với kỳ vọng bằng μ và độ lệch chuẩn bằng σ hữu hạn . Định lý giới hạn trung tâm sẽ cho chúng ta biết về dáng điệu tiệm cận của phân bố xác suất của tổng $S_n = X_1 + \ldots + X_n$, khi n tiến tới vô cùng. Trước khi xét dáng điệu tiệm cận của S_n, chúng ta sẽ *chuẩn hóa* nó. Bởi vì nếu để nguyên, và giả sử chẳng hạn $\mu > 0$, thì theo luật số lớn, phân bố xác suất của S_n sẽ bị dồn về phía vô

cùng khi n tiến tới vô cùng, và như vậy thì nó không thể tiến tới một phân bố cho trước nào đó. Nhắc hệ quả sau đây của sự độc lập của các biến X_i:

$$\mathbb{E}(S_n) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n\mu, \quad var(S_n) = \sum_{i=1}^{n} var(X_i) = n\sigma^2 \tag{4.8}$$

Đặt $Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$, ta có

$$\mathbb{E}(Z_n) = 0, \quad var(Z_n) = 1. \tag{4.9}$$

Điều đó có nghĩa là, qua phép biến đổi tuyến tính $Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$, ta có thể đưa biến ngẫu nhiên S_n về một biến ngẫu nhiên Z_n có kỳ vọng bằng 0 và phương sai bằng 1. Biến ngẫu nhiên Z_n này được gọi là **chuẩn hóa** của S_n, hay còn gọi là **tổng chuẩn hóa** của $X_1, \ldots, X_n$. Sau khi đã chuẩn hóa như vậy, ta có thể so sánh dáng điệu của phân bố của Z_n với các phân bố chuẩn hóa khác (có cùng kỳ vọng bằng 0 và độ lệch chuẩn bằng 1). Định lý giới hạn trung tâm phát biểu rằng, bất kể phân bố ban đầu (của X_1) ra sao, khi n lớn thì phân bố của tổng chuẩn hóa Z_n có thể được xấp xỉ rất tốt bằng phân bố normal $\mathcal{N}(0,1)$, và khi n tiến tới vô cùng thì nó tiến tới $\mathcal{N}(0,1)$. Nói một cách chính xác hơn:

Định lý 4.3 (Định lý giới hạn trung tâm). *Giả sử $X_1, X_2, \ldots, X_n, \ldots$ là một dãy các biến ngẫu nhiên độc lập có cùng phân bố xác suất với kỳ vọng bằng μ và độ lệch chuẩn bằng σ hữu hạn. Đặt*

$$Z_n = \frac{(\sum_{i=1}^{n} X_i) - n\mu}{\sigma\sqrt{n}}. \tag{4.10}$$

Khi đó với mọi $a, b \in \mathbb{R}$*,* $a < b$*, ta có:*

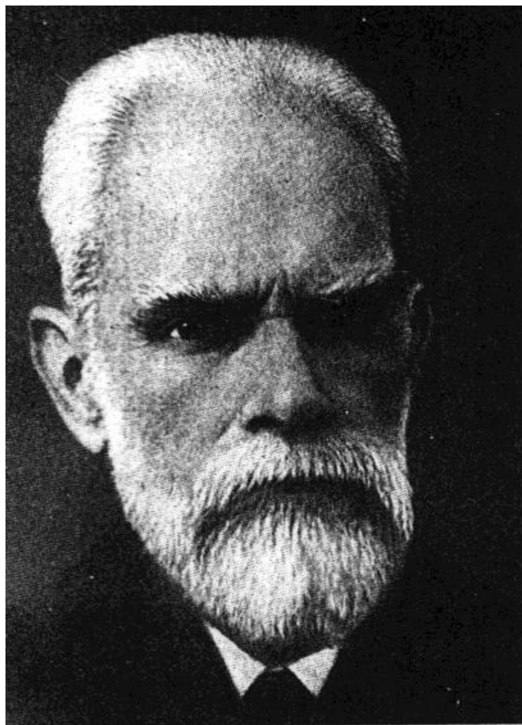
$$\lim_{n\to\infty} P\left(a \le Z_n \le b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \tag{4.11}$$

Một cách phát biểu tương đương là:

Định lý 4.4 (Định lý giới hạn trung tâm). *Giả sử* $X_1, X_2, \ldots, X_n, \ldots$ *là một dãy các biến ngẫu nhiên độc lập có cùng phân bố xác suất với kỳ vọng bằng* μ *và độ lệch chuẩn bằng* σ *hữu hạn. Đặt* $Z_n = \frac{(\sum_{i=1}^n X_i) - n\mu}{\sigma\sqrt{n}}$. *Khi đó với mọi tập con* $A \subset \mathbb{R}$ *thuộc sigma-đại số Borel, ta có:*

$$\lim_{n\to\infty} P_{Z_n}(A) = \lim_{n\to\infty} P\left(Z_n \in A\right) = \int_A \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = P_{\mathcal{N}(0,1)}(A). \tag{4.12}$$

Ghi chú 4.2. Nhiều nhà toán học đã đóng góp vào định lý giới hạn trung tâm: đầu tiên là de Moivre trong thế kỷ 18, rồi đến Laplace, Cauchy, Bessel, Poisson trong thế kỷ 19, rồi đến các nhà toán học Chebyschev, Markov, Lyapunov cuối 19 đầu thế kỷ 20, rồi đến các nhà toán học của thế kỷ 20 như von Mises, Polya, Lindeberg, Cramér phát triển và mở rộng nó, v.v. Tên gọi định lý giới hạn trung tâm (tiếng Đức: zentraler Grenzwertsatz) là do George Polya đưa ra năm 1920 trong một bài báo nhan đề như vậy. Một điều thú vị là Alan Turing (một trong những cha tổ của tin học hiện đại) cũng viết luận án về định lý giới hạn trung tâm vào năm 1934, trước khi phát hiện ra rằng kết quả của mình đã được Lindeberg làm ra từ năm 1922. Người đầu tiên phát biểu và chứng minh định lý giới hạn trung tâm cho một phân bố tổng quát có lẽ là Alexandr Mikhailovich Lyapunov (1857–1918), một nhà toán học người Nga, học trò của Chebyschev,



Александр Михайлович
ЛЯПУНОВ

Hình 4.2: Alexandr M. Lyapunov (1857–1918)

vào năm 1901. Ngoài công trình về xác suất, Lyapunov còn nổi tiếng về các công trình trong phương trình vi phân và sự ổn định của các hệ động lực (ổn định Lyapunov, các lũy thừa Lyapunov, v.v.).

Bài tập 4.2. Một nhà máy sản xuất dây xích bằng thép, mỗi dây gồm nhiều mắt xích. Độ dài của các mắt xích được định nghĩa sao cho độ

dài của dây xích bằng tổng độ dài các mắt xích. Phòng nghiên cứu của nhà máy đo thấy độ dài các mắt xích là một biến ngẫu nhiên X có kỳ vọng là 5cm và độ lệch chuẩn là 0,1cm. Nhà máy bán loại dây xích dài 50m, và để yên tâm về độ dài, xây xích đó được nối bằng 1002 mắt xích. Nhà máy cam đoan rằng không có dây xích nào loại này dài dưới 50m, và nếu khách hàng nào mua phải dây dài dưới 50m thì được đền tiền và được tặng một dây khác miễn phí.

i) Ước lượng xác suất để sao cho một dây xích với 1002 mắt xích có độ dài dưới 50m.

ii) Sau một thời gian, bộ phận bán hàng của nhà máy thấy có nhiều dây xích dài dưới 50m bị trả lại, và hỏi phòng nghiên cứu xem vấn đề nằm ở đâu. Sau khi điều tra, phòng nghiên cứu phát hiện là đo không thật chính xác: kỳ vọng của chiều dài mắt xích không phải là 5cm mà là 4,993cm. Với kỳ vọng này, xác suất để một dây xích với 1002 mắt xích có độ dài dưới 50m là bao nhiêu?

Bài tập 4.3. i) Chứng minh rằng tổng của n biến ngẫu nhiên độc lập có phân bố Poisson với tham số 1 là một biến ngẫu nhiên có phân bố Poisson với tham số n.

ii) Dùng kết quả trên và định lý giới hạn trung tâm để chứng minh khẳng định sau:

$$\lim_{n\to\infty} P(X_n \leq n) = 1/2,$$

trong đó X_n là biến ngẫu nhiên có phân bố Poisson với tham số n, và từ đó suy ra:

$$\lim_{n\to\infty} e^{-n}\left(1+\frac{n}{1!}+\frac{n^2}{2!}\ldots+\frac{n^n}{n!}\right)=\frac{1}{2}.$$

4.1.3 Giới hạn của dãy hàm đặc trưng

Để chứng minh định lý giới hạn trung tâm, chúng ta sẽ xét các hàm đặc trưng Φ_{Z_n} của các tổng chuẩn hóa $Z_n = \frac{(\sum_{i=1}^{n} X_i) - n\mu}{\sigma\sqrt{n}}$, trong đó $X_1, X_2, \ldots, X_n, \ldots$ là một dãy các biến ngẫu nhiên độc lập có cùng phân bố xác suất với kỳ vọng bằng μ và độ lệch chuẩn bằng σ hữu hạn.

Mệnh đề 4.5. *Với mọi $s \in \mathbb{R}$ ta có*

$$\lim_{n\to\infty} \Phi_{Z_n}(s) = \exp(-s^2/2). \qquad (4.13)$$

Chứng minh. Theo công thức biến đổi hàm đặc trưng khi biến đổi biến ngẫu nhiên một cách tuyến tính (xem khẳng định iii) của định lý 2.18), và công thức tính hàm đặc trưng của một tổng các biến ngẫu nhiên độc lập (xem khẳng định iv) của đinh lý 3.6), ta có:

$$\begin{aligned}\Phi_{Z_n}(s) &= \exp\left(\frac{-\sqrt{-1}n\mu}{\sigma\sqrt{n}}s\right)\Phi_{\sum_{i=1}^{n} X_i}\left(\frac{s}{\sigma\sqrt{n}}\right)\\ &= \exp\left(\frac{-\sqrt{-1}n\mu}{\sigma\sqrt{n}}s\right)\prod_{i=1}^{n}\Phi_{X_i}\left(\frac{s}{\sigma\sqrt{n}}\right)\\ &= \exp\left(\frac{-\sqrt{-1}n\mu}{\sigma\sqrt{n}}s\right)\left(\Phi_{X_1}\left(\frac{s}{\sigma\sqrt{n}}\right)\right)^n,\end{aligned}$$

do đó

$$\begin{aligned}\ln(\Phi_{Z_n}(s)) &= \frac{-\sqrt{-1}n\mu}{\sigma\sqrt{n}}s + n\ln\left(\Phi_{X_1}\left(\frac{s}{\sigma\sqrt{n}}\right)\right)\\ &= -\sqrt{-1}n\mu t + n\ln(\Phi_{X_1}(t)), \qquad (4.14)\end{aligned}$$

trong đó $t = \frac{s}{\sigma\sqrt{n}}$. Khi n tiến tới ∞ thì t tiến tới 0. Hàm $\Phi_{X_1}(t)$ khả vi liên tục 2 lần và có $\Phi_{X_1}(0) = 1$, $\Phi'_{X_1}(0) = \sqrt{-1}\mu$, $\Phi''_{X_1}(0) =$

$-\mathbb{E}(X_1^2) = -(\sigma^2 + \mu^2)$. (Xem định lý 2.18). Do đó hàm $\ln \Phi_{X_1}$ cũng khả vi liên tục hai lần trong lân cận của 0, và $\ln \Phi_{X_1}(0) = 0$, $(\ln \Phi_{X_1})'(0) = \sqrt{-1}\mu$, $(\ln \Phi_{X_1})''(0) = -\sigma^2$ Theo công thức khai triển Taylor–Lagrange, ta có:

$$\ln(\Phi_{X_1}(t)) = \sqrt{-1}\mu t - \frac{1}{2}\sigma^2 t^2 + o(t^2),$$

trong đó $o(t^2)$ là ký hiệu Landau: $o(t^2)/t^2$ tiến tới 0 khi t tiến tới 0. Do đó

$$\begin{aligned}\ln(\Phi_{Z_n}(s)) &= -\sqrt{-1}n\mu t + n(\sqrt{-1}\mu t - \frac{1}{2}\sigma^2 t^2 + o(t^2)) \\ &= -\frac{1}{2}n\sigma^2 t^2 + no(t^2) = -\frac{1}{2}s^2 + no(t^2).\end{aligned}$$

Khi n tiến tới vô cùng thì $no(t^2) = \frac{s^2}{\sigma^2}o(t^2)/t^2$ tiến tới 0, do đó

$$\begin{aligned}\lim_{n\to\infty} \Phi_{Z_n}(s) &= \exp(\lim_{n\to\infty} \ln(\Phi_{Z_n}(s))) \\ &= \exp(\lim_{n\to\infty} -\frac{1}{2}s^2 + no(t^2)) = \exp(-\frac{1}{2}s^2),\end{aligned}$$

là điều phải chứng minh. □

Nhắc lại rằng hàm $\Phi(s) = \exp(-s^2/2)$ chính là hàm đặc trưng của phân bố normal chuẩn tắc $\mathcal{N}(0,1)$. Định lý giới hạn trung tâm 4.3 suy ra trực tiếp từ Mệnh đề 4.5 và mệnh đề sau:

Mệnh đề 4.6. *Giả sử có một dãy biến ngẫu nhiên Z_n với các hàm đặc trưng Φ_{Z_n} tương ứng sao cho, với mọi $s \in \mathbb{R}$, $\Phi_{Z_n}(s)$ hội tụ đến $\Phi(s) = \exp(-s^2/2)$ khi n tiến tới vô cùng. Khi đó với mọi $a, b \in \mathbb{R}$, $a < b$, ta có*

$$\lim_{n\to\infty} P(a \le Z_n \le b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \tag{4.15}$$

Mệnh đề trên là một trường hợp riêng của *định lý liên tục* 4.11 về tiêu chuẩn *hội tụ yếu* của các phân bố xác suất, mà chúng ta sẽ bàn đến trong phần sau.

4.2 Hội tụ yếu và các kiểu hội tụ khác

4.2.1 Hội tụ yếu và hội tụ theo phân phối

Định nghĩa 4.1. *Một dãy phân bố xác suất P_n (hay một dãy hàm phân phối xác suất $\mathcal{F}_n$ tương ứng) được gọi là* **hội tụ yếu** *đến một phân bố xác suất P_∞ (hay đến một hàm phân phối xác suất $\mathcal{F}_\infty$ tương ứng) nếu chúng thỏa mãn điều kiện sau: Với mọi điểm liên tục $x \in \mathbb{R}$ của $\mathcal{F}_\infty$ (tức là $P_\infty(x) = 0$), ta có*

$$\lim_{n\to\infty} \mathcal{F}_n(x) = \mathcal{F}_\infty(x). \tag{4.16}$$

Chúng ta có thể ký hiệu sự hội tụ yếu như sau:

$$P_n \xrightarrow{w} P_\infty \ , \ \mathcal{F}_n \xrightarrow{w} \mathcal{F}_\infty. \tag{4.17}$$

Chữ w phía trên có nghĩa là yếu (*weak* tiếng Anh). Hội tụ yếu là kiểu hội tụ hay dùng nhất cho các thông kê xác suất. Bởi vậy khi ta viết $\lim_{n\to\infty} P_n = P_\infty$ ta sẽ hiểu đó là giới hạn yếu, tức là P_n hội tụ yếu đến P_∞. Ví dụ sau cho thấy vì sao, trong định nghĩa trên, ta chỉ yêu cầu $\lim_{n\to\infty} \mathcal{F}_n(x) = \mathcal{F}_\infty(x)$ khi x là điểm liên tục của $\mathcal{F}_\infty(x)$.

Ví dụ 4.2. Giả sử $(c_n)_{n\in\mathbb{N}}$ là một dãy số thực tiến tới một số thực c_∞ khi n tiến tới vô cùng. Giả sử thêm rằng $c_n > c$ với mọi n. Gọi P_n (hay P_∞) là phân bố xác suất của hằng số c_n (hay c_∞), tức là

phân bố xác suất rời rạc tập trung tại điểm c_n (hay c_∞): $P_n(c_n) = 1$ (hay $P_\infty(c_\infty) = 1$). Khi đó ta muốn nói một cách tự nhiên rằng P_n hội tụ đến P_∞ khi n tiến tới vô cùng. Tuy nhiên $\mathcal{F}_n(c_\infty) = P_n(]-\infty, c_\infty]) = 0$ với mọi n trong khi $\mathcal{F}_\infty(c_\infty) = 1$, và bởi vậy điều kiện $\lim_{n\to\infty} \mathcal{F}_n(x) = \mathcal{F}_\infty(x)$ không thỏa mãn tại điểm $x = c_\infty$ (là điểm gián đoạn của hàm phân phối xác suất $\mathcal{F}_\infty$). Tại các điểm $x \neq c_\infty$ thì điều kiện này được thỏa mãn. Bởi vậy, trong ví dụ này, để có được sự hội tụ của dãy phân bố $(P_n)_{n\in\mathbb{N}}$ đến P_∞, ta phải dùng hội tụ yếu, như được định nghĩa ở trên.

Các phân bố xác suất rời rạc có thể hội tụ yếu đến các phân bố xác suất liên tục, và ngược lại, các phân bố xác suất liên tục cũng có thể hội tụ yếu đến các phân bố xác suất rời rạc.

Ví dụ 4.3. i) Với mỗi $n \in \mathbb{N}$, gọi P_n là phân bố xác suất đều trên đoạn thẳng $[0, 1/n]$ (với hàm mật độ bằng n trên đoạn thẳng đó). Khi n tiến tới vô cùng, thì P_n hội tụ yếu đến phân bố rời rạc P_∞ tập trung tại điểm 0: $P_\infty(0) = 1$.
ii) Với mỗi $n \in \mathbb{N}$, gọi P_n là phân bố xác suất rời rạc tập trung tại n điểm $1/n, 2/n, \ldots, 1$ với các xác suất bằng nhau và bằng $1/n$: $P_n(1/n) = P_n(2/n) = \ldots = P_n(1) = 1/n$. Khi n tiến tới vô cùng, thì P_n hội tụ yếu đến phân bố đều trên đoạn thẳng $[0, 1]$.

Định nghĩa 4.2. *Một dãy biến ngẫu nhiên Z_n được gọi là* **hội tụ theo phân phối xác suất** *đến một biến ngẫu nhiên Z (hay còn gọi là hội tụ theo phân phối đến phân bố xác suất của Z), nếu như dãy phân bố xác suất P_{Z_n} của Z_n hội tụ yếu đến phân bố xác suất P_Z.*

Chúng ta sẽ ký hiệu sự hội tụ theo phân phối như sau:

$$Z_n \xrightarrow{d} Z, \tag{4.18}$$

hoặc là

$$Z_n \xrightarrow{d} P_Z. \tag{4.19}$$

Chữ d có nghĩa là *distribution*, tức là phân phối (hay phân bố) xác suất.

Ví dụ 4.4. Giả sử X_n là biết ngẫu nhiên rời rạc nhận hai giá trị $1/n$ và $1 - 1/n$ với các xác suất tương ứng $P(X_n = 1/n) = (n+1)/2n$ và $P(X_n = 1 - 1/n) = (n-1)/2n$. Khi đó X_n hội tụ theo phân phối đến biến ngẫu nhiên X với phân bố Bernoulli: $P(X = 0) = P(X = 1) = 1/2$.

Vì phân bố normal chuẩn tắc $\mathcal{N}(0, 1)$ là một phân phân bố liên tục, nên $Z_n \xrightarrow{d} \mathcal{N}(0, 1)$ khi và chỉ khi $\mathcal{F}_{Z_n}(b) \overset{n\to\infty}{\longrightarrow} \int_{\infty}^{b} \exp(-x^2/2)dx$ với mọi $x \in \mathbb{R}$. Bởi vậy định lý giới hạn trung tâm có thể được phát biểu lại như sau:

Định lý 4.7 (Định lý giới hạn trung tâm). *Giả sử $X_1, X_2, \ldots, X_n, \ldots$ là một dãy các biến ngẫu nhiên độc lập có cùng phân bố xác suất với kỳ vọng bằng μ và độ lệch chuẩn bằng σ hữu hạn. Gọi $Z_n = \dfrac{(\sum_{i=1}^{n} X_i) - n\mu}{\sigma\sqrt{n}}$ là tổng chuẩn hóa của $X_1, \ldots, X_n$. Khi đó*

$$Z_n \xrightarrow{d} \mathcal{N}(0, 1) \tag{4.20}$$

khi n tiến tới vô cùng.

Bài tập 4.4. Chứng minh rằng một dãy phân bố xác suất normal $\mathcal{N}(\mu_n, \sigma_n^2)$ hội tụ yếu khi và chỉ khi hai dãy số (μ_n) và σ_n hội tụ.

Bài tập 4.5. Chứng minh rằng một dãy phân bố xác suất P_n hội tụ yếu đến một phân bố xác suất P_∞ khi và chỉ khi với mọi đoạn thẳng mở $]a, b[$ ta có $\liminf_{n\to\infty} P_n(]a, b[) \geq P_\infty(]a, b[)$.

Bài tập 4.6. Giả sử rằng X_n có phân bố hình học với tham số $1/n$. Chứng minh rằng

$$\frac{X_n}{n} \xrightarrow{d} Y$$

khi n tiến tới vô cùng, trong đó Y có phân bố mũ với tham số 1.

Bài tập 4.7. Giả sử $X_1, X_2, \ldots$ là một dãy các biến ngẫu nhiên độc lập có phân bố đều $U(0, 1)$. Đặt

$$Y_n = n(1 - \max(X_1, X_2, \ldots, X_n)).$$

Chứng minh rằng Y_n hội tụ theo phân phối xác suất đến X, trong đó X có phân bố mũ với tham số 1.

4.2.2 Các metric trên không gian các phân bố xác suất

Về mặt trực giác, khi chúng ta nói rằng phân bố xác suất P_1 gần bằng phân bố xác suất P_2 có nghĩa là khoảng cách giữa P_1 và P_2 nhỏ. Nhưng để phát biểu điều đó một cách chính xác, ta cần định nghĩa khoảng cách ở đây là gì. Có nhiều cách định nghĩa khác nhau, cho ra các kết quả khác nhau, trên không gian các phân bố xác suất. Ở đây chúng ta sẽ bàn đến 3 cách trong số các cách định nghĩa.

Định nghĩa 4.3. *Giả sử P_X và P_Y là hai phân bố xác suất trên $\mathbb{R}$, với các hàm phân phối xác suất tương ứng $\mathcal{F}_X$ và $\mathcal{F}_Y$.*

i) **Khoảng cách** L_1 *(với hạch nhân $e^{-|x|}$) giữa P_X và P_Y là đại lượng*

$$d_1(P_X, P_Y) = \int_{-\infty}^{\infty} |\mathcal{F}_X(x) - \mathcal{F}_Y(x)| e^{-|x|} dx. \tag{4.21}$$

ii) **Khoảng cách Lévy–Prokhorov** *giữa P_X và P_Y là đại lượng*

$$d_{LP}(P_X, P_Y) = \inf\{\varepsilon > 0 \mid P_X(A) \leq P_Y(A^\varepsilon) + \varepsilon \text{ và } P_Y(A) \leq P_X(A^\varepsilon) + \varepsilon \; \forall \; A \in \mathcal{B}(\mathbb{R})\}, \quad (4.22)$$

trong đó $\mathcal{B}(\mathbb{R})$ là đại số Borel trên $\mathbb{R}$, và

$$A^\varepsilon = \{x \in \mathbb{R} \mid \exists y \in A \text{ sao cho } |x - y| < \varepsilon\} \quad (4.23)$$

là $\varepsilon-$lân cận của A trong $\mathbb{R}$.
iii) **Khoảng cách Kolmogorov–Smirnov** *giữa P_X và P_Y là đại lượng*

$$d_{KS}(P_X, P_Y) = \sup_{x \in \mathbb{R}} |\mathcal{F}_X(x) - \mathcal{F}_Y(x)|. \quad (4.24)$$

Nhắc lại rằng, một **metric** trên một không gian V là một ánh xạ $d : V \times V \to \mathbb{R}$ thỏa mãn các tính chất sau:
i) Dương tính: $d(u, v) \geq 0$ với mọi $u, v \in V$, và $d(u, v) = 0$ khi và chỉ khi $u = v$.
ii) Đối xứng: $d(u, v) = d(v, u)$ với mọi $u, v \in V$.
iii) Bất đẳng thức tam giác: $d(u, v) + d(v, w) \geq d(u, w)$ với mọi $u, v, w \in V$.
Một không gian V với một metric d trên đó được gọi là một **không gian metric**, và $d(u, v)$ được gọi là **khoảng cách** giữa u và v (theo metric d). Một không gian với một metric d trên đó thì trở thành một không gian tôpô, trong đó sự hội tụ của một dãy điểm $(u_n)_{n \in \mathbb{N}}$ đến một điểm u_∞ (theo metric d) có nghĩa là $d(u_n, u_\infty)$ tiến tới 0 khi n tiến tới vô cùng.

Dễ dàng kiểm tra rằng, cả ba định nghĩa khoảng cách d_1, d_{LP} và d_{KS} phía trên đều thỏa mãn các tiên đề của một metric, do đó ta có

3 metric khác nhau trên không gian các phân bố xác suất trên $\mathbb{R}$, ứng với 3 định nghĩa khoảng cách này. Quan hệ giữa 3 metric d_1, d_{LP} và d_{KS} như sau:

Định lý 4.8. *Hai metric d_1 và d_{LP} tương đương với nhau về tôpô (cho cùng một tôpô trên không gian các phân bố xác suất trên $\mathbb{R}$), nghĩa là $\lim_{n\to\infty} d_1(P_n, P_\infty) = 0$ khi và chỉ khi $\lim_{n\to\infty} d_{LP}(P_n, P_\infty) = 0$. Metric mạnh d_{KS} mạnh hơn hai metric d_1 và d_{LP}, nghĩa là nếu $\lim_{n\to\infty} d_{KS}(P_n, P_\infty) = 0$ thì $\lim_{n\to\infty} d_1(P_n, P_\infty) = 0$ và $\lim_{n\to\infty} d_{LP}(P_n, P_\infty) = 0$, nhưng khẳng định ngược lại không đúng.*

Khẳng định d_{KS} mạnh hơn d_1 khá là hiển nhiên: dễ dàng thấy rằng

$$d_1(P_1, P_2) \leq d_{KS}(P_1, P_2).\int_{-\infty}^{\infty} e^{-|x|}dx = 2d_{KS}(P_1, P_2)$$

với hai phân bố xác suất P_1, P_2 bất kỳ trên $\mathbb{R}$. Dãy phân bố xác suất trong ví dụ 4.2 cho thấy d_{KS} thực sự mạnh hơn d_1, tức là có thể có $d_1(P_n, P_\infty)$ tiến tới 0 trong khi $d_{KS}(P_n, P_\infty)$ không tiến tới 0 khi n tiến tới vô cùng. Sự tương đương tôpô của d_1 và d_{LP} là một bài tập giải tích thú vị dành cho bạn đọc. Định nghĩa d_1 đơn giản hơn định nghĩa d_{LP}. Nhưng lợi thế của d_{LP} nằm ở tính tổng quát của nó: nó dùng được cho không gian các phân bố xác suất trên một không gian metric bất kỳ. Chú ý thêm rằng, hàm $e^{-|x|}$ trong định nghĩa metric d_1 được chọn một cách khá tùy tiện. Nếu ta thay hàm đó bằng một hàm khác, thoả mãn các tính chất bị chặn liên tục dương có tích phân trên $\mathbb{R}$ hữu hạn, thì ta được một định nghĩa metric khác, tương đương về mặt tô pô với metric d_1.

Định lý 4.9. *Một dãy các phân bố xác suất P_n hội tụ theo metric d_1 (hay là metric d_{LP}) đến một phân bố xác suất P_∞ (có nghĩa là $\lim_{n\to\infty} d_1(P_n, P_\infty) = 0$) khi và chỉ khi P_n hội tụ yếu P_∞ khi n tiến tới vô cùng. Nói cách khác, sự hội tụ yếu trùng với sự hội tụ theo metric d_1 và trùng với sự hội tụ theo metric Lévy-Prokhorov.*

Chứng minh.

i)Điều kiện cần. Giả sử có một điểm liên tục x_0 của $\mathcal{F}_\infty$ sao cho $\mathcal{F}_n(x_0)$ không hội tụ đến $\mathcal{F}_\infty(x_0)$. Khi đó tồn tại một hằng số $c > 0$ và một dãy số tự nhiên n_k tiến tới vô cùng sao cho $|\mathcal{F}_{n_k}(x_0) - \mathcal{F}_\infty(x_0)| > c$ với mọi $k \in \mathbb{N}$. Ta sẽ giả sử $\mathcal{F}_\infty(x_0) - \mathcal{F}_{n_k}(x_0) > c$ với mọi k. (Trường hợp có thể chọn $\mathcal{F}_{n_k}(x_0) - \mathcal{F}_\infty(x_0) > c$ với mọi k hoàn toàn tương tự). Do tính liên tục của $\mathcal{F}_\infty$ tại x_0, tồn tại $\delta > 0$ đủ nhỏ sao cho $\mathcal{F}_\infty(x_0) - \mathcal{F}_\infty(x) < c/2$ với mọi $x \in [x_0 - \delta, x_0]$. Do các hàm phân bố xác suất là hàm tịnh tiến tăng, ta có $\mathcal{F}_{n_k}(x) \le \mathcal{F}_{n_k}(x_0)$ với mọi $x \in [x_0 - \delta, x_0]$, từ đó suy ra $\mathcal{F}_\infty(x) - \mathcal{F}_{n_k}(x) > c/2$ với mọi $x \in [x_0 - \delta, x_0]$, và do đó tồn tại một hằng số dương $C = \int_{x_0-\delta}^{x_0} (c/2) e^{-|x|} dx > 0$, sao cho $d(P_{n_k}, P_\infty) > C$ với mọi k. Điều đó có nghĩa là $d(P_n, P_\infty)$ không tiến tới 0 khi n tiến tới vô cùng.

ii) Điều kiện đủ. Giả sử $\mathcal{F}_\infty(x) = \lim_{n\to\infty} \mathcal{F}_n(x)$ tại mọi điểm liên tục của $\mathcal{F}_\infty$. Giả sử $\epsilon > 0$ là một số dương tùy ý. Nhắc lại rằng hàm $\mathcal{F}_\infty$ là một hàm đơn điệu không giảm bị chặn, và tập các điểm không liên tục của $\mathcal{F}_\infty$ là hữu hạn hoặc đếm được. Ta có thể chọn một dãy hữu hạn $x_0 < x_1 < \ldots < x_N$ các điểm liên tục của $\mathcal{F}_\infty$ sao cho $\int_{-\infty}^{x_0} e^{-|x|} dx < \epsilon$, $\int_{x_N}^{\infty} e^{-|x|} dx < \epsilon$, và với mọi $k = 0, 1, \ldots, N - 1$ ta có hoặc là $0 \le \mathcal{F}_\infty(x_{k+1}) - \mathcal{F}_\infty(x_k) < \epsilon/(x_N - x_0)$ hoặc là $0 < x_{k+1} - x_k < \epsilon/N$. Gọi I là tập các chỉ số k thỏa mãn $0 \le$

$\mathcal{F}_\infty(x_{k+1}) - \mathcal{F}_\infty(x_k) < \epsilon/(x_N - x_0)$, và đặt $J = \{0, \ldots, N-1\} \setminus I$. Do dãy hàm số $\mathcal{F}_n$ tiến tới $\mathcal{F}_\infty$ tại các điểm $x_0, \ldots, x_N$, tồn tại một số tự nhiên K sao cho với mọi $n \geq K$ và mọi $i = 0, 1, \ldots, N$ ta có $|\mathcal{F}_n(x_i) - \mathcal{F}_\infty(x_i)| < \epsilon/(x_N - x_0)$. Nếu $k \in I$ thì từ các bất đẳng thức này cùng với bất đẳng thức $0 \leq \mathcal{F}_\infty(x_{k+1}) - \mathcal{F}_\infty(x_k) < \epsilon/(x_N - x_0)$ và sự đơn điệu không giảm của $\mathcal{F}_n$ và $\mathcal{F}_\infty$ suy ra bất đẳng thức sau:

$$|\mathcal{F}_n(x) - \mathcal{F}_\infty(x)| < 2\epsilon/(x_N - x_0)\ \forall x \in [x_k, x_{k+1}]$$

(với mọi $k \in I$). Ta chia $d(P_n, P_\infty)$, với mọi $n \geq K$, thành 3 phần:

$$d(P_n, P_\infty) = \int_{-\infty}^{\infty} |\mathcal{F}_n(x) - \mathcal{F}_\infty(x)| e^{-|x|} dx = A_n + B_n + C_n,$$

với

$$\begin{aligned}
A_n = &\int_{-\infty}^{x_0} |\mathcal{F}_n(x) - \mathcal{F}_\infty(x)| e^{-|x|} dx + \\
&\textstyle\int_{x_N}^{\infty} |\mathcal{F}_n(x) - \mathcal{F}_\infty(x)| e^{-|x|} dx < 2\epsilon, \\
B_n = &\sum_{k \in I} \int_{x_i}^{x_{i+1}} |\mathcal{F}_n(x) - \mathcal{F}_\infty(x)| e^{-|x|} dx < \\
&< \textstyle\sum_{k \in I} \int_{x_i}^{x_{i+1}} \frac{2\epsilon}{x_N - x_0} dx \leq \int_{x_0}^{x_N} \frac{2\epsilon}{x_N - x_0} dx = 2\epsilon, \\
C_n = &\sum_{k \in J} \int_{x_i}^{x_{i+1}} |\mathcal{F}_n(x) - \mathcal{F}_\infty(x)| e^{-|x|} dx \leq \\
&\leq \textstyle\sum_{k \in J} (x_{k+1} - x_k) \leq N \max_{k \in J} (x_{k+1} - x_k) < N \frac{\epsilon}{N} = \epsilon.
\end{aligned}$$

Tổng cộng lại, ta có $d(P_n, P_\infty) < 5\epsilon$, với mọi n đủ lớn. Vì ϵ là tùy ý, nên $\lim\limits_{n \to \infty} d(P_n, P_\infty) = 0$. □

Bài tập 4.8. Chứng minh sự tương đương về tôpô của metric d_1 và metric d_{LP}.

4.2.3 Định lý tiền compact của Prokhorov

Định nghĩa 4.4. *Một dãy các phân bố xác suất $(P_n)_{n\in\mathbb{N}}$ được gọi là* **chặt** *(tight) nếu như với mọi $\epsilon > 0$ tồn tại $R_\epsilon \in \mathbb{R}_+$ sao cho*

$$P_n([-R_\epsilon, R_\epsilon]) > 1 - \epsilon \;\forall\; n \in \mathbb{N}. \tag{4.25}$$

Nói một cách nôm na, điều kiện *chặt* là điều kiện "xác suất không bị dàn trải về vô cùng" khi n tiến tới vô cùng.

Định lý 4.10 (Prokhorov). *Giả sử $(P_n)_{n\in\mathbb{N}}$ là một dãy phân bố xác suất trên $\mathbb{R}$ thỏa mãn điều kiện chặt. Khi đó tồn tại một dãy con $(P_{k_n})_{n\in\mathbb{N}}$ ($k_n \to \infty$ khi $n \to \infty$) hội tụ yếu đến một phân bố xác suất nào đó.*

Tính chất "mọi dãy điểm (của một tập nào đó) đều có một dãy con hội tụ" gọi là tính chất *tiền compact* (pre-compact). Bởi vậy định lý trên của Prokhorov được gọi là *định lý tiền compact*.

Sơ lược chứng minh. Lấy một tập trù mật đếm được trên $\mathbb{R}$ (ví dụ như tập hợp $\mathbb{Q}$ các số hữu tỷ), và đánh số thứ tự các số trong tập đó thành một dãy số $(a_m)_{m\in\mathbb{N}}$. Có thể xây dựng bằng qui nạp theo m một dãy con $(P_{k_n})_{n\in\mathbb{N}}$ của dãy phân bố xác suất $(P_n)_{n\in\mathbb{N}}$ thỏa mãn tính chất sau: dãy số $\mathcal{F}_{k_n}(a_m)$ hội tụ với mọi $m \in \mathbb{N}$, trong đó $\mathcal{F}_{k_n}$ là các hàm phân phối xác suất tương ứng. Xây dựng hàm $\mathcal{F}_\infty$ như sau: Đặt $Q(a_m) = \lim_{n\to\infty} \mathcal{F}_{k_n}(a_m)$, và $\mathcal{F}_\infty(x) = \inf\{Q(a_m) | a_m > x\}$ với mọi $x \in \mathbb{R}$. Dễ thấy hàm $\mathcal{F}_\infty$ thỏa mãn các tính chất đơn điệu không giảm và liên tục bên phải. Tính chất chặt của dãy $(P_{k_n})_{n\in\mathbb{N}}$ đảm bảo rằng $\mathcal{F}_\infty(x)$ tiến tới 0 khi x tiến tới $-\infty$ và tiến tới 1 khi x tiến tới $+\infty$. Bởi vậy nó là hàm phân phối của một phân bố xác suất P_∞ nào

đó. Bước cuối cùng là kiểm tra rằng P_{k_n} hội tụ yếu tới P_∞. (Bài tập: Làm chi tiết các bước chứng minh). $\square$

Ghi chú 4.3. Định lý Prokhorov và metric Lévy-Prokhorov là gọi theo tên của Yuri Vasilevich Prokhorov (sinh năm 1929), một nhà toán học Nga Xô Viết chuyên về xác suất, học trò của Kolmogorov, viện sĩ viện hàn lâm khoa học Liên Xô từ năm 1972 (nay là viện hàn lâm khoa học Nga).

4.2.4 Định lý liên tục

Định lý 4.11 (Định lý liên tục). *Giả sử có một phân bố xác suất P_∞ và một dãy phân bố xác suất P_n trên $\mathbb{R}$. Khi đó ba điều kiện sau đây tương đương với nhau:*
1) Dãy phân bố xác suất P_n hội tụ yếu đến P_∞ khi n tiến đến vô cùng.
2) Với mọi hàm liên tục và bị chặn F trên $\mathbb{R}$ ta có

$$\lim_{n\to\infty} \int_{\mathbb{R}} F dP_n = \int_{\mathbb{R}} F dP_\infty. \tag{4.26}$$

3) Gọi Φ_n và Φ_∞ là các hàm đặc trưng tương ứng của P_n và P_∞. Khi đó với mọi $s \in \mathbb{R}$ ta có

$$\lim_{n\to\infty} \Phi_n(s) = \Phi_\infty(s). \tag{4.27}$$

Chứng minh. Điều kiện 1) suy ra điều kiện 2): Giả sử điều kiện 1) được thỏa mãn, và giả sử F là một hàm liên tục bị chặn: tồn tại một số thực dương M sao cho $|F(x)| \leq M$ với mọi $x \in \mathbb{R}$. Gọi $\epsilon > 0$ là một số dương bất kỳ. Chúng ta sẽ chứng minh rằng

$$\left| \int_{\mathbb{R}} F dP_n - \int_{\mathbb{R}} F dP_\infty \right| < \epsilon \tag{4.28}$$

với mọi n đủ lớn. Tồn tại $R \in \mathbb{R}_+$ sao cho $-R$ và R là hai điểm liên tục của $\mathcal{F}_\infty$, và

$$\mathcal{F}_\infty(-R) < \epsilon/6M \ , \ \mathcal{F}_\infty(R) > 1 - \epsilon/6M. \tag{4.29}$$

Khi đó, với mọi n đủ lớn, ta cũng có $\mathcal{F}_n(-R) < \epsilon/6M$ và $\mathcal{F}_n(R) > 1 - \epsilon/6M$. Vì giá trị tuyệt đối của F bị chặn bởi M, nên từ đó ta có

$$\left| \int_{]-\infty,-R]} F dP_\infty \right| < \epsilon/6 \ , \ \left| \int_{]R,+\infty[} F dP \infty \right| < \epsilon/6, \tag{4.30}$$

và

$$\left| \int_{]-\infty,-R]} F dP_n \right| < \epsilon/6 \ , \ \left| \int_{]R,+\infty[} F dP_n \right| < \epsilon/6 \tag{4.31}$$

với mọi n đủ lớn. Như vậy, để chứng minh bất đẳng thức 4.28, ta chỉ cần chứng minh rằng

$$\left| \int_{]-R,R]} F dP_n - \int_{]-R,R]} F dP_\infty \right| < \epsilon/3 \tag{4.32}$$

với mọi n đủ lớn. Vì hàm F liên tục, nên nó liên tục đều trên đoạn thẳng $[-R, R]$. Bởi vậy tồn tại một dãy số $a_0 = -R < a_1 < \ldots < a_N = R$, sao cho các số a_i đều là các điểm liên tục của $\mathcal{F}_\infty$, và trên mỗi đoạn thẳng $[a_{i-1}, a_i]$ độ dao động của F nhỏ hơn $\epsilon/6$: $|F(x) -$

$F(a_i)| < \epsilon/9$ với mọi $x \in [a_{i-1}, a_i]$. Từ đó suy ra

$$\begin{aligned}\left|\int_{]-R,R]} FdP_n - \sum_{i=1}^{N} F(a_i)(\mathcal{F}_n(a_i) - \mathcal{F}_n(a_{i-1}))\right| \\ = \left|\sum_{i=1}^{N}\int_{]a_{i-1},a_i]}(F - F(a_i))dP_n\right| \\ \leq \sum_{i=1}^{N}\int_{]a_{i-1},a_i]}|F - F(a_i)|dP_n < \sum_{i=1}^{N}\int_{]a_{i-1},a_i]}(\epsilon/9)dP_n \\ = (\epsilon/9)\int_{]-R,R]} 1dP_n \leq \epsilon/9 \end{aligned} \quad (4.33)$$

với mọi n, và một bất đẳng thức như vậy cho P_∞. Chú ý rằng các điểm a_i là các điểm liên tục của $\mathcal{F}_\infty$, do đó $\mathcal{F}_n(a_i)$ tiến tới $\mathcal{F}_\infty(a_i)$ khi n tiến tới vô cùng với mọi $i = 1, \ldots, N$. Bởi vậy với mọi n đủ lớn ta có

$$\left|\sum_{i=1}^{N} F(a_i)(\mathcal{F}_n(a_i) - \mathcal{F}_n(a_{i-1})) - \sum_{i=1}^{N} F(a_i)(\mathcal{F}_\infty(a_i) - \mathcal{F}_\infty(a_{i-1}))\right| < \epsilon/9. \quad (4.34)$$

Kết hợp các bất đẳng thức trên lại với nhau, ta được điều phải chứng minh .

Điều kiện 2) suy ra điều kiện 3): Điều kiện 3) chẳng qua là trường hợp riêng của điều kiện 2) cho các hàm số $F_s(x) = \exp(\sqrt{-1}sx)$, bởi vì, theo định nghĩa,

$$\Phi_X(s) = \mathbb{E}(\exp(\sqrt{-1}sX)) = \int_{\mathbb{R}} \exp(\sqrt{-1}sx)dP_X \quad (4.35)$$

với mọi phân bố xác suất P_X (với một biến ngẫu nhiên X tương ứng).

Điều kiện 3) suy ra điều kiện 1): (Sơ lược chứng minh). Giả sử $\lim_{n\to\infty}\Phi_n(s) = \Phi_\infty(s)$ với mọi $s \in \mathbb{R}$. Nhắc lại rằng các hàm đặc

trưng của các phân bố xác suất là bị chặn bởi 1: $|\Phi_n(s)| \leq 1$ với mọi $s \in \mathbb{R}$. Bởi vậy, theo định lý hội tụ bị chặn Lebesgue, ta có

$$\lim_{n\to\infty} \int_{-\epsilon}^{\epsilon} \Phi_n(s)ds = \int_{-\epsilon}^{\epsilon} \Phi_\infty(s)ds \tag{4.36}$$

với mọi $\epsilon > 0$. Mặt khác, ta có bất đẳng thức sau:

Bổ đề 4.12. *Với mọi biến ngẫu nhiên X, và mọi số $\epsilon > 0$ ta có:*

$$P_X([-\frac{2}{\epsilon}, \frac{2}{\epsilon}]) \geq \left|\frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon} \Phi_X(s)ds\right| - 1. \tag{4.37}$$

Chú ý rằng vế phải của bất đẳng thức (4.37) tiến tới 1 khi ϵ tiến tới 0. Từ bất đẳng thức này và công thức giới hạn (4.36) dễ dàng suy ra rằng dãy phân bố xác suất (P_n) thỏa mãn điều kiện chặt. Bởi vậy, theo định lý tiền compact của Prokhorov, tồn tại một dãy con P_{k_n} hội tụ yếu đến một phân bố xác suất $\tilde{P}$ nào đó. Như đã chứng minh ở trên, khi P_{k_n} hội tụ đến $\tilde{P}$, thì Φ_{k_n} cũng hội tụ đến hàm đặc trưng của $\tilde{P}$ tại mọi điểm. Thế nhưng Φ_n hội tụ đến Φ_∞, bởi vậy hàm đặt trưng của $\tilde{P}$ chính là Φ_∞. Vì mọi phân bố xác suất được xác định duy nhất bằng hàm đặc trưng của nó, nên $\tilde{P}$ chính là P_∞. Có nghĩa là có một dãy con của (P_n) hội tụ yếu đến P_∞. Nhưng khi đó, toàn bộ dãy (P_n) phải hội tụ yếu đến P_∞, vì nếu không, tương tự như trên, sử dụng định lý Prokhorov, ta sẽ tìm được một dãy con của (P_n) hội tụ yếu đến một phân bố xác suất $\hat{P}$ khác P_∞, nhưng $\hat{P}$ lại có hàm đặc trưng trùng với hàm đặc trưng của P_∞, là điều không thể xảy ra. □

Ghi chú 4.4. Định lý phía trên được gọi là *định lý liên tục*, vì nó khẳng định rằng ánh xạ từ các hàm đặc trưng vào các phân bố xác suất tương ứng là một ánh xạ liên tục. Nó là một phần của định lý liên

tục của Paul Pièrre Lévy (1886-1971), một nhà toán học người Pháp. Lévy là người đưa ra nhiều khái niệm quan trọng trong lý thuyết xác suất, trong đó có khái niệm *martingale*. Định lý liên tục của Lévy phát biểu như sau:

Định lý 4.13 (Lévy). *Giả sử các hàm đặc trưng Φ_{X_n} của các biến ngẫu nhiên X_n ($n \in \mathbb{N}$) tiến tới một hàm Φ tại mọi điểm trên $\mathbb{R}$ (hội tụ theo từng điểm). Khi đó các khẳng định sau đây là tương đương:*
i) X_n hội tụ theo phân phối xác suất đến một biến ngẫu nhiên X nào đó.
ii) Dãy các phân bố xác suất $(P_{X_n})_{n\in\mathbb{N}}$ thỏa mãn điều kiện chặt.
iii) Φ là hàm đặc trưng của một biến ngẫu nhiê X nào đó.
iv) Φ là một hàm liên tục trên $\mathbb{R}$.
v) Hàm $\Phi(s)$ liên tục tại điểm $s = 0$.

Bài tập 4.9. (Chứng minh bổ đề 4.12). Chứng minh đẳng thức sau

$$\begin{aligned}\frac{1}{2\epsilon}\int_{-\epsilon}^{\epsilon}\Phi_X(s)ds &= \int_{x\in\mathbb{R}}\frac{\sin(\epsilon x)}{\epsilon x}dP_X \\ &= \int_{|x|\leq 2/\epsilon}\frac{\sin(\epsilon x)}{\epsilon x}dP_X + \int_{|x|>2/\epsilon}\frac{\sin(\epsilon x)}{\epsilon x}dP_X \quad (4.38)\end{aligned}$$

với mọi biến ngẫu nhiên X. (Gợi ý: dùng định nghĩa của hàm đặc trưng, và công thức thay đổi thứ tự tính tính phân Fubini). Sau đó áp dụng các bất đẳng thứ $|\sin(t)/t| \leq 1$ với mọi $t \in \mathbb{R}$ và $|\sin(t)/t| \leq 1/2$ với mọi $|t| \geq 2$, $t \in \mathbb{R}$ vào đẳng thức trên, để suy ra bất đẳng thức (4.37).

4.2.5 Các kiểu hội tụ khác của dãy biến ngẫu nhiên

Ngoài hội tụ theo phân phối (là kiểu hội tụ trong định lý giới hạn trung tâm), chúng ta đã gặp những kiểu hội tụ sau đây: hội tụ theo xác suất (là kiểu hội tụ trong dạng yếu của luật số lớn), và hội tụ hầu như chắc chắn (là kiểu hội tụ trong dạng mạnh của luật số lớn)

Định nghĩa 4.5. *Một dãy biến ngẫu nhiên X_n được gọi là* **hội tụ theo xác suất** *đến một biến ngẫu nhiên X nếu như với mọi $\epsilon > 0$ ta có*

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0. \tag{4.39}$$

Định nghĩa 4.6. *Một dãy biến ngẫu nhiên X_n được gọi là* **hội tụ hầu như chắc chắn** *đến một biến ngẫu nhiên X nếu như*

$$P(\{\omega \in \Omega | \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1, \tag{4.40}$$

trong đó Ω ký hiệu không gian xác suất chung của các biến ngẫu nhiên X_n và X, và ω ký hiệu các phần tử của Ω, tức là các sự kiện thành phần.

Sự hội tụ hầu như chắc chắn còn được gọi là sự **hội tụ hầu khắp mọi nơi**.

Ngoài ra, có một loại hội tụ khác hay được dùng đến, là hội tụ theo chuẩn L_k ($k \geq 1$ không nhất thiết phải là số nguyên; trường hợp hay dùng nhất là $k = 2$):

Định nghĩa 4.7. *Đại lượng $(\mathbb{E}(|X|^k))^{1/k}$ được gọi là* **chuẩn** *L_k của một biến ngẫu nhiên X. Một dãy biến ngẫu nhiên X_n được gọi là* **hội**

tụ theo chuẩn L_k *(hay còn gọi là* **hội tụ theo trung bình cấp** k*) đến một biến ngẫu nhiên* X *nếu như*

$$\lim_{n\to\infty} \mathbb{E}(|X_n - X|^k) = 0. \quad (4.41)$$

Định lý 4.14 (Quan hệ giữa các kiểu hội tụ)**.** *i) Nếu* $k_1 > k_2 \geq 1$*, thì sự hội tụ theo chuẩn* L_{k_1} *mạnh hơn sự hội tụ theo chuẩn* L_{k_2}*. Có nghĩa là, nếu* X_n *hội tụ theo chuẩn* L_{k_1} *thì nó cũng hội tụ theo chuẩn* L_{k_2}*. (Điều ngược lại nói chung không đúng).*
ii) Với mọi $k \geq 1$*, sự hội tụ theo chuẩn* L_k *mạn hơn sự hội tụ theo xác suất.*
iii) Sự hội tụ hầu như chắc chắn mạnh hơn sự hội tụ theo xác suất.
iv) Sự hội tụ theo xác suất mạnh hơn sự hội tụ theo phân phối.

Ghi chú 4.5. Sự hội tụ theo chuẩn L_k không suy ra sự hội tụ hầu như chắc chắn, và ngược lại sự hội tụ hầu như chắc chắn cũng không mạnh hơn sự hội tụ theo chuẩn L_k.

4.3 Phân bố χ^2 và định lý Pearson

Phân bố **ki bình phương** (χ^2, chi-square) với tham số $r \in \mathbb{N}$ là phân bố xác suất của biến ngẫu nhiên χ^2_r định nghĩa như sau:

$$\chi^2_r = Z_1^2 + \ldots Z_r^2, \quad (4.42)$$

trong đó $Z_1, \ldots, Z_r$ là một bộ r biến ngẫu nhiên độc lập tuân theo phân bố normal chuẩn tắc $\mathcal{N}(0,1)$. Tham số r ở đây được gọi là **số bậc tự do**. Chẳng hạn khi $r = 3$ thì người ta nói là có 3 bậc tự do (3 degrees of freedom).

Phân bố χ^2 hay xuất hiện trong những bài toán thống kê, mà chúng ta sẽ xét đến ở chương sau. Nó liên quan đến việc ước lượng phương sai của một phân bố xác suất normal. Đồng thời, nó đóng vai trò rất quan trọng trong việc kiểm định các giả thuyết về dáng điệu các phân bố xác suất, qua cái gọi là **kiểm định** χ^2 (chi-square test). Cơ sở của kiểm định χ^2 là định lý giới hạn sau đây của Karl Pearson:

Định lý 4.15 (Pearson). *Giả sử X là một biến ngẫu nhiên nhận hữu hạn các giá trị $x_1, \ldots, x_s$ với các xác suất $P_X(x_i) = p_i > 0$ tương ứng ($\sum_{i=1}^s p_i = 1$). Giả sử $X_1, X_2, \ldots, X_n, \ldots$ là một dãy các biến ngẫu nhiên độc lập và có cùng phân phối xác suất với X. Với mỗi n, gọi $\nu_i = \nu_{i,n}$ là biến ngẫu nhiên sau: ν_i là số lần xuất hiện giá trị x_i trong dãy $X_1, \ldots, X_n$. ($\sum_{i=1}^s \nu_i = n$). Khi đó*

$$\sum_{i=1}^{k} \frac{(\nu_i - np_i)^2}{np_i} \xrightarrow{d} \chi^2_{s-1}, \tag{4.43}$$

tức là biến ngẫu nhiên $\displaystyle\sum_{i=1}^{s} \frac{(\nu_i - np_i)^2}{np_i}$ hội tụ theo phân phối đến χ^2_{s-1}, khi n tiến tới vô cùng.

Ghi chú 4.6. Trong trường hợp $s = 2$, và để cho tiện giả sử $x_1 = 1$, $x_2 = 0$ (các giá trị của x_i không quan trọng, chỉ có xác suất của chúng là quan trọng trong định lý Pearson), ta có: X tuân theo phân bố Bernoulli với tham số $p = p_1 = \mathbb{E}(X)$, $1 - p = p_2$, $\nu_1 = \sum_{i=1}^n X_i$, $\nu_2 = n - \nu_1$, và

$$\begin{aligned}\frac{(\nu_1 - np_1)^2}{np_1} + \frac{(\nu_2 - np_2)^2}{np_2} &= \frac{(\nu_1 - np)^2}{pn} + \frac{(\nu_1 - np)^2}{(1-p)n} \\ &= \frac{(\nu_1 - np)^2}{p(1-p)n} = \left(\frac{\sum_{i=1}^n X_i - n\mathbb{E}(X)}{\sqrt{n}\sigma(X)}\right)^2.\end{aligned}$$

Theo định lý giới hạn trung tâm thì $\frac{\sum_{i=1}^n X_i - n\mathbb{E}(X)}{\sqrt{n}\sigma(X)} \xrightarrow{d} Z_1$ (trong đó Z_1 có phân bố normal chuẩn tắc $\mathcal{N}(0,1)$), và do đó

$$\left(\frac{\sum_{i=1}^n X_i - n\mathbb{E}(X)}{\sqrt{n}\sigma(X)}\right)^2 \xrightarrow{d} \chi_1^2.$$

Nói cách khác, định lý Pearson trong trường hợp $k = 2$ là hệ quả trực tiếp của định lý giới hạn trung tâm. Trong trường hợp $k > 2$, định lý Pearson có thể coi như một mở rộng của định lý giới hạn trung tâm.

Sơ lược chứng minh định lý Pearson. Đặt $F_i = \frac{(\nu_i - np_i)}{\sqrt{p_i(1-p_i)n}}$. Ta cần tìm giới hạn theo phân phối xác suất của $\sum_{i=1}^s (1-p_i)F_i^2$, khi n tiến tới vô cùng. Theo định lý giới hạn trung tâm, ta có $F_i \xrightarrow{d} \mathcal{N}(0,1)$ với mọi $i = 1, \ldots, s$ khi n tiến tới vô cùng. Tuy nhiên, các biến $F_1, \ldots, F_s$ có phụ thuộc vào nhau: $\sum_{i=1}^s \sqrt{p_i(1-p_i)}F_i = 0$. Bằng cách tính trực tiếp, ta có: $cov(F_i, F_j) = -\sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}}$ với mọi $i \neq j$.

Một điểm đáng chú ý là, cũng theo định lý giới hạn trung tâm, với mọi $c_1, \ldots, c_s \in \mathbb{R}$, $\sum_{i=1}^s c_i F_i$ cũng hội tụ theo phân phối đến một phân bố normal. Từ đó suy ra vector ngẫu nhiên $(F_1, \ldots, F_s)$ hội tụ theo phân phối đến một vector ngẫu nhiên $(G_1, \ldots, G_s)$ với phân bố normal nhiều chiều $\mathcal{N}(0, \Sigma)$, trong đó ma trận hiệp phương sai Σ được xác định như sau:

$$\Sigma_{ii} = var(G_i) = 1 \text{ và } \Sigma_{ij} = cov(G_i, G_j) = -\sqrt{p_i p_j/((1-p_i)(1-p_j))}. \tag{4.44}$$

với mọi i, j. Điều còn lại cần phải chứng minh là $\sum_{i=1}^s (1-p_i)G_i^2$ có cùng phân bố xác suất với χ_{s-1}^2.

Ma trận hiệp phương sai Σ suy biến (bởi vì $\sum_{i=1}^{s}\sqrt{p_i(1-p_i)}G_i = 0$), có hạng (rank) bằng $s-1$, do đó (về mặt phân phối xác suất) ta có thể nhận được vector ngẫu nhiên $(G_1,\ldots,G_s)$ từ một vector ngẫu nhiên $(Z_1,\ldots,Z_{s-1})$ có phân bố normal chuẩn tắc $(s-1)$ chiều, qua một phép biến đổi tuyến tính:

$$(G_1,\ldots,G_s)^t = (a_{ij})_{i=1,\ldots,s}^{j=1,\ldots,s-1}.(Z_1,\ldots,Z_{s-1})^t. \tag{4.45}$$

Theo định lý 3.23, ta cần chọn ma trận A sao cho $A.A^t = \Sigma$. Ma trận $A = (a_{ij})_{i=1,\ldots,s}^{j=1,\ldots,s-1}$ có thể được chọn như sau. Gọi $O = (o_{ij})_{i=1,\ldots,s}^{j=1,\ldots,s}$ là một ma trận vuông góc (orthogonal, có nghĩa là $O.O^t = \mathbb{I}_s$) bất kỳ thỏa mãn điều kiện: $o_{si} = \sqrt{p_i}$ với mọi $i = 1,\ldots,s$, tức là cột cuối cùng của O được cho bởi các số $\sqrt{p_i}$. Ma trận vuông góc O như vậy tồn tại bởi vì $\sum_{i=1}^{s}(\sqrt{p_i})^2 = 1$. Đặt

$$a_{ij} = \frac{o_{ij}}{\sqrt{1-p_i}} \text{ với mọi } i = 1,\ldots,s,\ j = 1,\ldots,s-1. \tag{4.46}$$

Dễ dàng kiểm tra rằng ma trận A định nghĩa như trên thỏa mãn điều kiện $A.A^t = \Sigma$, và như vậy ta có thể coi rằng $(G_1,\ldots,G_s)^t = (a_{ij})_{i=1,\ldots,s}^{j=1,\ldots,s-1}.(Z_1,\ldots,Z_{s-1})^t$. Nói cách khác, ta có

$$G_i = \sum_{j=1}^{s-1}\frac{o_{ij}}{\sqrt{1-p_i}}Z_j \text{ với mọi } i = 1,\ldots,s, \tag{4.47}$$

từ đó suy ra:

$$\sum_{i=1}^{s}(1-p_i)G_i^2 = \sum_{i=1}^{s}(1-p_i)\left(\sum_{j=1}^{s-1}\frac{o_{ij}}{\sqrt{1-p_i}}Z_j\right)^2$$
$$= \sum_{i=1}^{s}\left(\sum_{j=1}^{s-1}o_{ij}^2Z_j^2 + \sum_{j\neq k}o_{ij}o_{ik}Z_jZ_k\right)$$
$$= \sum_{j=1}^{s-1}(\sum_{i=1}^{s}o_{ij}^2)Z_j^2 + \sum_{j\neq k}(\sum_{i=1}^{s}o_{ik})Z_jZ_k = \sum_{j=1}^{s-1}Z_j^2 \quad (4.48)$$

và ta được điều phải chứng minh. □

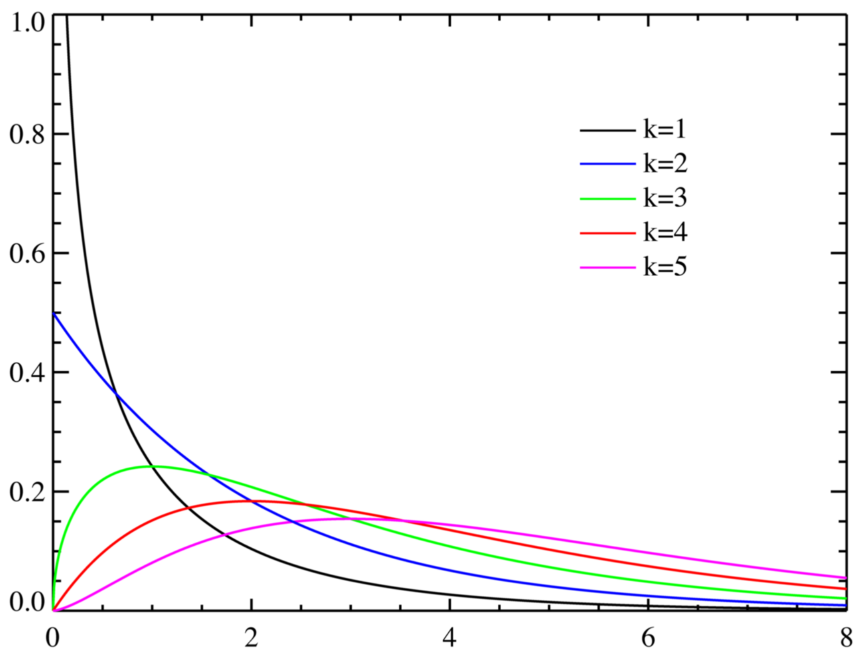
Do tầm qua trọng của phân bố χ^2 trong thống kê, nên nó được nghiên cứu rất kỹ, và có thể tính hàm phân phối của nó bằng máy tính hoặc tra bảng. Hàm mật độ của phân bố χ^2 là hàm sau:

Định lý 4.16. *Phân bố χ^2 với r bậc tự do $(r > 0)$ có hàm mật độ là:*

$$\rho(x) = \begin{cases} \dfrac{1}{2^{r/2}\Gamma(r/2)}x^{\frac{r}{2}-1}e^{-\frac{x}{2}} & \textit{khi } x > 0 \\ 0 & \textit{khi } x \leq 0 \end{cases}, \quad (4.49)$$

trong đó Γ là hàm gamma: $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx$.

Ghi chú 4.7. Karl Pearson (1857–1936), người Anh, được coi là một trong những cha tổ của ngành thống kê toán học. Năm 33 tuổi, sau khi đọc sách *Natural Inheritance* của Francis Galton, Pearson bắt đầu quan tâm đến các phương pháp thống kê, để áp dụng chúng vào việc kiểm nghiệm học thuyết sàng lọc tự nhiên của Darwin, trong khuôn khổ của học thuyết *eugenics* (ưu sinh học) đang thịnh hành thời đó, mà Pearson là một trong những người đi theo. Pearson là người lập



Hình 4.3: Hàm mật độ của χ_k^2, với $k = 1, 2, 3, 4, 5$

ra khoa thống kê đầu tiên, năm 1911, tại University College London. Nhiều khái niệm cơ bản trong xác suất thống kê là dựa trên những công trình của Pearson, trong đó có: hệ số tương quan, hồi qui tuyến tính, phân loại các phân bố xác suất, kiểm định ki bình phương.

Bài tập 4.10. Làm chi tiết các bước trong chứng minh của định lý Pearson 4.15.



Hình 4.4: Karl Pearson

Chương 5

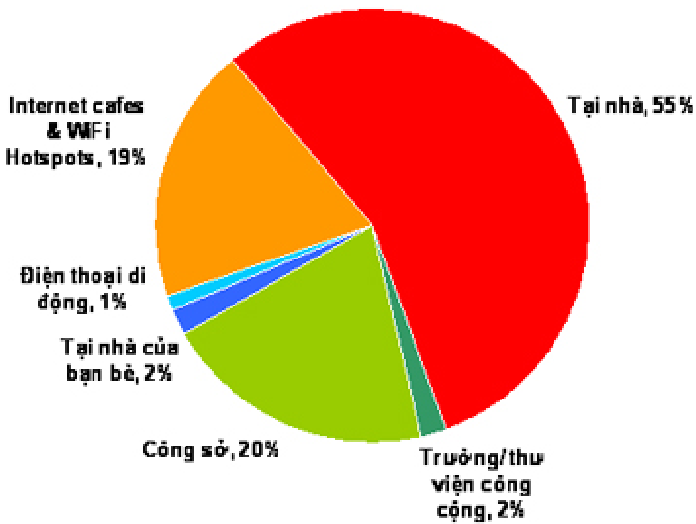
Thống kê toán học

5.1 Các vấn đề thống kê

Thống kế toán học có thể coi là tổng thể các phương pháp toán học, dựa trên lý thuyết xác suất và các công cụ khác, nhằm đưa ra được những thông tin mới, kết luận mới, có giá trị, từ những bảng số liệu thô ban đầu, và nhằm giải quyết những vấn đề nào đó nảy sinh từ thực tế. Có thể kể tên một số mục đích chính của thống kê như sau:
- Mổ tả số liệu.
- Ước lượng và dự đoán các đại lượng.
- Tìm ra các mối quan hệ giữa các đại lượng .
- Kiểm định các giả thuyết.

Thống kê học là một ngành lớn, với nhiều phương pháp khác nhau để dùng cho các tình huống khác nhau (có người ví các phương

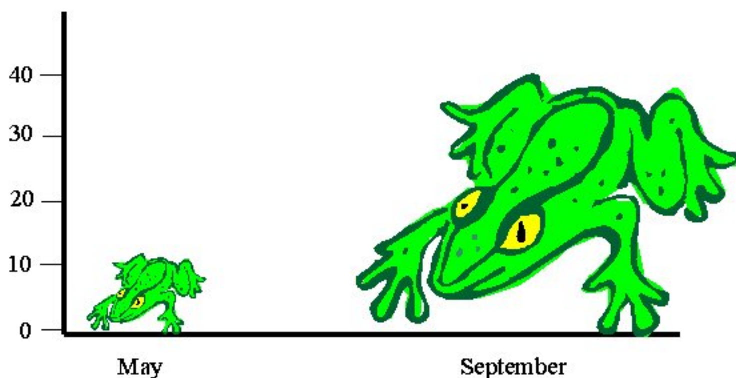
pháp thông kê như là các cách nấu ăn, rất đa dạng phong phú), và có nhiều điểm cần chú ý để khỏi dẫn đến các kết luận thống kê sai lệch (hoặc là bị mắc lừa bởi những người cố tình làm thống kê theo các phương pháp sai lệch). Trong chương này chúng ta sẽ chỉ bàn tới một số vấn đề và phương pháp thống kê toán học cơ bản nhất. Trước khi đi vào lý thuyết, ở phần này chúng ta sẽ điểm qua các mục đích chính trên của thống kê, qua một số ví dụ.



Hình 5.1: Tỷ lệ thời gian dùng internet ở Việt Nam năm 2009

Ví dụ 5.1. (*Biểu đồ thống kê*). Trong thống kê mô tả, ngoài các bảng số liệu (cùng với một số đại lượng đặc trưng tiêu biểu nhất như trung vị, kỳ vọng, phương sai), các biểu đồ cũng hay được dùng, để giúp

người đọc nắm bắt thông tin về số liệu một cách nhanh chóng. Một số loại biểu đồ hay gặp là: biểu đồ tần số, đồ thị phát tán, biểu đồ hình quạt (pie chart), v.v. Hình 5.1, là một ví dụ về biểu đồ hình quạt, phản ảnh tỷ lệ thời gian dùng internet ở Việt Nam vào năm 2009 (theo báo Lao Động). So với các bảng số liệu, các biểu đồ có thể có nhược điểm là cho thông tin không được chính xác bằng (độ sai số cao hơn), nhưng có ưu điểm là cho được cùng một lúc nhiều thông tin trên một hình ảnh, dễ tiếp thu đối với não người hơn là một bảng các con số.



Hình 5.2: Số ếch trong hồ vào tháng 5 và tháng 9

Tất nhiên, có những biểu đồ có thể phản ánh rất sai lệch về các đại lượng. Hình 5.2 là một ví dụ đơn giản về đề tài *nói dối bằng thống kê*. Đồ thị đó xuất phát từ số liệu thống kê số ếch trong 1 cái hồ: hơn 10 con vào tháng 5, và nhiều gấp 3 lần như vậy vào tháng 9. Nhưng nhìn vào đồ thị người ta sẽ có cảm giác là số ếch vào tháng 9 gấp

$3 \times 3 = 9$ lần tháng 5.

Ví dụ 5.2. (*Phát xít Đức sản xuất bao nhiêu máy bay và xe cơ giới?*). Trong chiến tranh, việc ước lượng được đúng sức mạnh của quân địch là một việc nhiều khi có tính chất sống còn. Trong chiến tranh thế giới lần thứ II, các cơ quan tình báo quân đồng minh Anh-Mỹ đã cung cấp nhiều thông tin rất sai lệch về lực lượng quân Đức. Thế nhưng, bằng phương pháp thống kê (thu nhặt các mã số trên các xác máy bay, lốp xe, v.v. của quân Đức bị bắn cháy, bỏ rơi, rồi từ đó giải mã và dùng các hàm ước lượng để ước lượng), nhà thống kê học Richard Ruggles cùng với các cộng sự của mình, lúc đó làm tại Cục tình báo kinh tế của Anh, đã ước lượng được rất chính xác số máy bay và xe cơ giới mà Đức sản xuất được hàng tháng.(1)

Công suất hàng tháng của Phát xít Đức	Máy bay	Xe cơ giới
Ước lượng của Ruggles	28500	147000
Số liệu thực theo tài liệu của Đức	26400	159000

Trong khi đó, ước lượng của các tình báo viên Anh-Mỹ là công suất của Đức khoảng 1 triệu xe cơ giới một tháng!

Ví dụ 5.3. (*Thần dược chống béo phì?*). Tỷ lệ số người bị béo phì (obesity) tăng rất nhanh trên thế giới (kể cả ở Việt Nam, châu Âu, và Mỹ) trong những thập kỷ cuối thế kỷ 20 - đầu thế kỷ 21, và trở thành một vấn đề xã hội lớn, vì béo phì hay dẫn đến nhiều căn bệnh

(1)Theo sách [1], dựa trên: Ruggles, R. and H. Brodie, "An Empirical Approach to Economic Intelligence in World War II", Journal of the American Statistical Association, 42, March 1947; và theo: James Tobin, "In memoriam: Richard Ruggles (1916–2001)" Review of Income and Wealth Series 47, Number 3, September 2001

khác (tim mạch, tiểu đường, đột quị, vô sinh, v.v.), và có thể làm giảm đáng kể tuổi thọ của người. Chống béo phì là một vấn đề nóng hổi, nhưng cho đến năm 2009 chưa có thuốc nào thật hiệu quả được bán trên thị trường. Điều này có thể thay đổi trong những năm sau đó, vì trong năm 2009 có 3 hãng dược phẩm công bố các kết quả thử nghiệm lâm sàng giai đoạn III (phase III clinical trial) cho các loại thuốc chống béo phì mới có nhiều triển vọng. Trong đó đáng chú ý nhất có lẽ là thuốc Qnexa của hãng Vivus. Công bố kết quả về Qnexa của Vivus vào ngày 09/09/2009[2] có một bảng thống kê sau (trong số nhiều bảng thống kê):

	ITT-LOCF			Completers		
EQUIP (OB-302) 56 Weeks	Placebo (n=498)	Qnexa Low Dose (n=234)	Qnexa Full Dose (n=498)	Placebo (n=241)	Qnexa Low Dose (n=138)	Qnexa Full Dose (n=301)
Mean Weight Loss (%)	1.6%	5.1%*	11.0%*	2.5%	7.0%*	14.7%*
Greater than or equal to 5% weight loss rate	17%	45%*	67%*	26%	59%*	84%*

ITT-LOCF: Intent-to-treat with last observation carried forward
*p<0.0001 vs. placebo

Theo bảng trên, tổng số người tham gia thử nghiệm lâm sàng (trong thử nghiệm đó) là $498 + 234 + 498 = 1300$ người. Đợt thử nghiệm kéo dài 56 tuần, nhưng có những người bỏ dở giữa chừng: trong số 498 người được nhận placebo (trông giống như viên thuốc thật, nhưng

[2]Nguồn: http://ir.vivus.com/releasedetail.cfm?ReleaseID=407933

không có thuốc trong đó) thì chỉ có 241 người theo đến cùng cuộc thử nghiệm, còn trong số 498 người được nhận liều đầy đủ của thuốc, có 301 người (61%) theo đến cùng. Trong số những người được nhận đủ liều và theo đến cùng, thì có 84% số người giảm được ít nhất 5% trọng lượng, và trung bình mỗi người giảm được 14,7% trọng lượng.

Trong bảng trên có viết $p < 0,0001$ vs. placebo. Điều đó có nghĩa là, với độ tin cậy bằng $1 - p > 99,99\%$ (hay nói cách khác, với khả năng kết luận sai lầm nhỏ hơn $0,01\%$), các con số thống kê cho thấy kết quả đạt được (ở đây là giảm cân) tốt hơn khi có thuốc so với khi không có thuốc. Thông thường, khi $p < 0,01$ thì người ta chấp nhận giả thuyết là thuốc có hiệu ứng thực sự, còn nếu $p \geq 0,05$ thì hiệu ứng đó không rõ ràng, có thể là do ngẫu nhiên.

Các hãng dược phẩm trên thế giới, trước khi được quyền bán một loại thuốc mới nào đó, thông thường đều phải qua thử nghiệm lâm sàng diện rộng (trên ít nhất mấy trăm bệnh nhân), và các kết quả thống kê phải chứng tỏ rõ ràng công dụng và sự an toàn của thuốc, tức là phải qua được kiểm định thống kê cho giả thuyết "thuốc có công dụng và an toàn", với độ tin cậy cao.

Ví dụ 5.4. (*London nguy hiểm hay an toàn?*[(3)]). Ngày 10/07/2008, có 4 vụ giết người bằng dao ở 4 nơi khác nhau ở London. Sự kiện này làm náo loạn dư luận đến mức thủ tướng Anh là Gordon Brown phải tuyên bố hứa sẽ tìm cách làm giảm các vụ đâm dao. London có trở nên nguy hiểm cho tính mạng hơn những năm trước không? Để trả lời câu hỏi đó, chúng ta có thể dựa trên một số số liệu thống kê sau:

- Trong 5 năm trước đó, mỗi năm ở London có khoảng 170 người bị

[(3)] Dựa theo tạp chí Significance của Royal Statistical Society, số tháng 3/2009

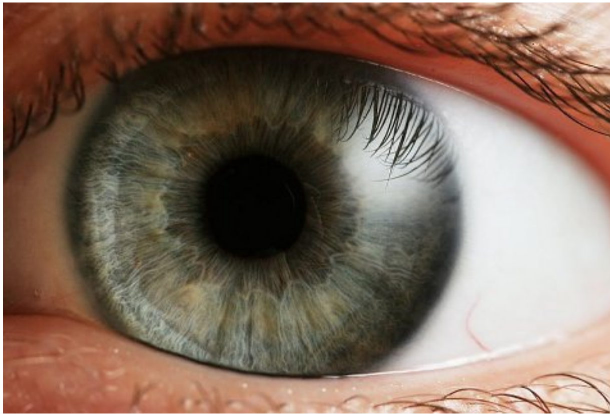
giết, và con số này khá ổn định hàng năm.
- Khoảng 41% các vụ giết người là dùng dao, 17% là dùng súng, 9% là đánh đập (không vũ khí), 5% là đánh bằng vật không phải dao, 3% là bóp cổ, 3% là dùng thuốc độc, v.v., và 17% là không xác định được phương pháp.
- Trong thời gian 3 năm 04/2004 – 03/2007, có 713 ngày không có vụ án mạng nào, 299 ngày có 1 vụ, 66 ngày có 2 vụ, 16 ngày có 3 vụ, 1 ngày có 4 vụ, và không có ngày nào có từ 5 vụ trở lên.

Từ các số liệu thống kê, người ta tính được một số ước lượng sau về số vụ án mạng ở London:

- Số vụ án mạng xảy ra trong ngày tuân theo phân bố Poisson với kỳ vọng là 0,44 (tức là trung bình mỗi ngày có 0,44 vụ) .
- Kỳ vọng là mỗi năm có khoảng 3-4 ngày với 3 vụ án mạng, cứ khoảng gần 3 năm thì có 1 ngày với 4 vụ án mạng, và khoảng 30 năm thì mới có một ngày với 5 vụ án mạng.

Việc xảy ra 1 ngày vào năm 2008 với 4 vụ án mạng không nằm ngoài các con số ước lượng trên. Đâm bằng dao là phương pháp gây án mạng phổ biến nhất (41% tổng số các vụ). Khi có 4 vụ án mạng, thì xác suất để cả 4 vụ đều do đâm dao là $(0,41)^4 = 2,8\%$, một con số khá nhỏ, nhưng cũng không nhỏ đến mức "không thể xảy ra". Khi có 4 vụ án mạng xảy ra cùng ngày, thì có rất nhiều tổ hợp các khả năng xảy ra về phương pháp gây án mạng trong 4 vụ đó (ví dụ 2 vụ dùng dao, 1 vụ dùng súng, 1 vụ thắt cổ), và tất cả các tổ hợp đó đều có xác suất nhỏ, tổ hợp với xác suất lớn nhất cũng không vượt quá 6%. Từ đó, có thể kết luận là, việc hôm 10/07/2008 xảy ra 4 án mạng ở London, và cả 4 đều bằng đâm dao, hoàn toàn nằm trong các ước

lượng về án mạng xảy ra ở London, và không hề chứng tỏ xu thế gì mới. Tổng kết năm 2008, ở London có 152 án mạng xảy ra năm đó. Phương tiện truyền thông được dịp vui mừng vì "đã lâu rồi chưa năm nào London được an toàn như vậy". Nhưng con số đó có chứng tỏ xu thế gì không, hay chẳng qua cũng chỉ là một sự ngẫu nhiên không nằm ngoài qui luật chung?



Hình 5.3: Các đường vân trong màng mắt

Ví dụ 5.5. (*Con mắt trở thành chìa khóa*). Đầu thế kỷ 21, đã có những khách sạn mà khách không cần chìa khóa phòng, chỉ cần nhìn vào camera ở cửa phòng, là phòng tự động mở cửa. Sự tiện lợi này dựa trên công nghệ nhận biết danh tính của người qua màng mắt (iris). Một điều thú vị là, kể cả khi hai người sinh đôi và trông giống hệt nhau, thì các đường nét trong màng mắt của họ vẫn rất khác nhau, do quá trình phát triển các đường nét trong màng mắt ở thai nhi phụ thuộc vào nhiều yếu tố ngẫu nhiên (không do di truyền). Từ những

năm 1930, các bác sĩ mắt đã nói rằng có thể dùng màng mắt để nhận biết danh tính người. John Daugman là một trong những người làm ra công nghệ nhận biết danh tính bằng màng mắt, từ cuối thế kỷ 20. Thuật toán của ông ta tách ra được từ ảnh màng mắt 1 mã với 266 đơn vị thông tin có thể coi là ngẫu nhiên và độc lập với nhau (mỗi đơn đơn vị ở đây là một biến ngẫu nhiên nhận 2 giá trị 0 và 1, với xác suất $50\% - 50\%$, và các biến này gần như độc lập với nhau). Để tìm ra 266 đơn vị thông tin độc lập đó (xuất phát từ 2048 đơn vị thông tin không độc lập với nhau) và kiểm định sự độc lập của chúng, Daugman đã làm thống kê so sánh hơn 222 nghìn lần cặp ảnh màng mắt khác chủ (2 mắt trong 1 cặp là của hai người khác nhau), và hơn 500 cặp ảnh màng mắt cùng chủ[(4)]. Một trong các kết quả là, tỷ lệ đơn vị thông tin chệch nhau giữa mã của 2 mắt khác chủ tuân theo phân bố normal với kỳ vọng là 45.6% (tức là trung bình hai mắt khác chủ thì có 45.6% đơn vị thông tin chệch nhau) với độ lệch chuẩn là 0.18%, và không có cặp mắt khác chủ nào (trong các thử nghiệm) có dưới 37% đơn vị thông tin lệch nhau. Mặt khác, hai ảnh màng mắt khác nhau của cùng một chủ thì trung bình chỉ có 9% các đơn vị thông tin bị lệch nhau trong số 266 đơn vị, và không có cặp ảnh mắt cùng chủ nào bị lệch nhau quá 31% đơn vị thông tin. Từ đó dẫn đến thuật toán phân biệt: coi rằng nếu hai mã bị lệch nhau không quá 34% số đơn vị thông tin, thì vẫn là của cùng một người, còn nếu trên 34% thì coi là của hai người khác nhau.

[(4)] J. Daugman, Wavelet demodulation codes, statistical independence, and pattern recognition, in: Proceedings IMA-IP: Mathematical Methods, Algorithms, and Applications, (Blackledge and Turner eds.), Horwood, London, 2000, pages 244–260.

Một điều cần chú ý là, thống kê hay bị các tổ chức hay cá nhân lạm dụng để bóp méo sự thật theo hướng có lợi cho mình, hoặc có khi tự dối mình, nếu như làm không đúng cách. Có rất nhiều cách nói dối khác nhau bằng thống kê, chẳng hạn như: bịa đặt các con số không có thật, lựa chọn các con số có lợi, giấu đi các con số bất lợi, thiên lệch (bias) trong việc chọn mẫu thí nghiệm, v.v. Ví dụ về nói giối trắng trợn: Bộ quốc phòng Mỹ có tuyên bố rằng, trong cuộc chiến với Irak năm 1991, các tên lửa Patriot của Mỹ đã bắn rụng 41 tên lửa Scud của Irak, nhưng khi Quốc hội Mỹ điều tra lại thấy chỉ có 4 tên lửa Scud bị bắn rụng. Ví dụ về bias làm hỏng kết quả thống kê: Báo *Literacy Digest* thăm dò ý kiến cử tri về bầu cử tổng thống ở Mỹ năm 1936, qua điện thoại và qua các độc giả đặt báo. Kết quả thăm dò trên phạm vi rất rộng cho dự đoán là Landon sẽ được 370 phiếu (đại cử tri) còn Roosevelt sẽ chỉ được 161 phiếu. Thế nhưng lúc bầu thật thì Roosevelt thắng. Hoá ra, đối tượng mà Literacy Digest thăm dò năm đó, những người có tiền đặt điện thoại hay đặt báo, là những người thuộc tầng lớp khá giả, có bias theo phía Landon (Đảng Cộng hòa), không đặc trưng cho toàn dân chúng Mỹ.

Nói chung, để thống kê toán học cho ra được các kết quả đáng tin cậy, ngoài các công thức toán học đúng đắn, còn cần đảm bảo sự trung thực của các số liệu, có mẫu thực nghiệm (lượng số liệu) đủ lớn, và loại đi được ảnh hưởng của các bias để đảm bảo tính ngẫu nhiên của số liệu. Nhiều khi việc loại đi các kết quả có bias cao từ mẫu thực nghiệm là công việc hiệu quả, cho ra kết luận thống kê chính xác và đỡ tốn kém hơn là tăng cỡ của mẫu thực nghiệm lên thêm nhiều. Ở chương này, chúng ta sẽ chỉ bàn đến một số phương

pháp thống kê cơ bản, dựa trên giả sử là số liệu mà chúng ta nhận được là đúng thực và không bị bias.

5.2 Ước lượng bằng thống kê

5.2.1 Mẫu thực nghiệm và phân bố thực nghiệm

Chúng ta thử hình dung một tình huống sau: Một nhà sản xuất dưa chuột muối đóng hộp muốn biết phân bố chiều dài các quả dưa chuột (chiều dài trung bình, độ lệch chuẩn, ...), để làm vỏ hộp với kích thước thích hợp. Nhà sản xuất này sẽ không đi đo hết chiều dài của hàng triệu quả dưa chuột sẽ được đóng hộp. Họ sẽ chỉ đo chiều dài của một số n quả dưa chuột được chọn một cách ngẫu nhiên, rồi từ đó ước lượng ra phân bố chiều dài. Số n ở đây có thể là một con số khá lớn, ví dụ 100 quả hay 1000 quả, nhưng nó là một phần rất nhỏ của tổng số các quả dưa chuột.

Để mô hình hóa bài toán ước lượng trên, ta sẽ gọi X là biến ngẫu nhiên "chiều dài của quả dưa chuột". Chúng ta muốn ước lượng phân bố xác suất P_X của X, hoặc là ước lượng những đại lượng đặc trưng của P_X, ví dụ như kỳ vọng và phương sai. Để ước lượng, chúng ta sẽ lấy ra n giá trị của X một cách ngẫu nhiên (chọn ra n quả dưa chuột một cách ngẫu nhiên rồi đo chiều dài của chúng). Gọi các giá trị được lấy ra là $x_1, \ldots, x_n$. Bộ n giá trị $(x_1, \ldots, x_n)$ được gọi là một *mẫu thực nghiệm cỡ* n của biến ngẫu nhiên X

Nói một cách tổng quát, một **mẫu thực nghiệm** (empirical sample) **cỡ** n của một biến ngẫu nhiên X là một giá trị $\mathbf{x} = (x_1, \ldots, x_n)$

của vector ngẫu nhiên $\mathbf{X} = (X_1, \ldots, X_n)$, trong đó các biến ngẫu nhiên $X_1, \ldots, X_n$ độc lập và có cùng phân bố xác suất với X. (Trong ví dụ, X_i là biến ngẫu nhiên "chiều dài của quả dưa chuột thứ i được chọn", còn x_i là giá trị nhận được của X_i). Các số x_i được gọi là các **giá trị thực nghiệm** của X (hay của X_i).

Ghi chú 5.1. Trong thực tế, có những tình huống mà các biến ngẫu nhiên $X_1, \ldots, X_n$ không thể độc lập với nhau. Ví dụ, nếu gọi X_i là mã số của cái xác máy bay thứ i của phát xít Đức mà quân đồng minh nhặt được, thì X_i không thể bằng X_j khi $i \neq j$ và do đó X_i không độc lập với X_j. Trong những trường hợp như vậy, hoặc là sự phụ thuộc tuy có nhưng nhỏ, có thể bỏ qua, hoặc là ta phải điều chỉnh lý thuyết sau cho thích hợp. Ở đây, để đơn giản, ta sẽ luôn giả sử rằng các biến X_i độc lập với nhau.

Mẫu thực nghiệm $(x_1, \ldots, x_n)$ cho ta một phân bố xác suất $\hat{P}_n$ trên $\mathbb{R}$, gọi là **phân bố xác suất thực nghiệm**, như sau: nó là phân bố xác suất rời rạc tập trung tại các điểm $x_1, \ldots, x_n$, sao cho mỗi điểm x_i có tỷ trọng xác suất là $1/n$. Nói cách khác, nếu x_i khác tất cả các số còn lại thì $\hat{P}_n(x_i) = 1/n$. Nhưng nếu có k số bằng nhau, $x_{i_1} = x_{i_2} = \ldots = x_{i_k}$ và khác các số còn lại, thì $\hat{P}_n(x_{i_1}) = k/n$. Một cách định nghĩa khác của phân bố $\hat{P}_n$ này là qua **hàm phân phối thực nghiệm** $\hat{\mathcal{F}}_n$ của nó: với mọi $x \in \mathbb{R}$, $\hat{\mathcal{F}}_n(x)$ bằng $1/n$ nhân với số lượng các số x_i nhỏ hơn hoặc bằng x. Khẳng định sau là hệ quả trực tiếp của luật số lớn:

Định lý 5.1. *Hầu như chắc chắn rằng $\hat{P}_n$ hội tụ yếu đến P_X khi n tiến tới vô cùng.*

Nói cách khác, tập hợp các dãy vô hạn giá trị thực nghiệm

$$(x_1, x_2, \ldots, x_n, \ldots)$$

sao cho dãy các phân bố xác suất thực nghiệm $(\hat{P}_n)_{n\in\mathbb{N}}$ tương ứng không hội tụ yếu đến phân bố xác suất X_P của X là một tập có độ đo bằng 0. Không gian xác suất ở đây là tích vô hạn $\prod_{i=1}^{\infty}(\mathbb{R}, P_{X_i}) \cong (\mathbb{R}, P_X)^{\mathbb{N}}$, tương tự như trong phát biểu của dạng mạnh của luật số lớn.

Định lý 5.1 cho ta một nguyên tắc sau đây về ước lượng:

Phân bố xác suất của một biến ngẫu nhiên X có thể được ước lượng bằng các phân bố thực nghiệm của X, và khi cỡ của mẫu thực nghiệm càng cao thì ước lượng này càng chính xác. Các đại lượng đặc trưng của X có thể được ước lượng bằng các đại lượng đặc trưng tương ứng của các phân bố thực nghiệm.

Ví dụ 5.6. Kỳ vọng của phân bố thực nghiệm $\hat{P}_n$ của mẫu thực nghiệm $(x_1, \ldots, x_n)$ là

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{5.1}$$

Giá trị $\hat{\mu}$ được gọi là một **kỳ vọng thực nghiệm** (hay **kỳ vọng mẫu**) của X, và là một ước lượng của kỳ vọng của X. Tương tự như vậy, với mọi $k \in \mathbb{N}$, giá trị

$$\frac{1}{n}\sum_{i=1}^{n} x_i^k, \tag{5.2}$$

gọi là **moment thực nghiệm bậc** k, là một ước lượng của moment bậc k của X. Giá trị

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i)^2, \tag{5.3}$$

gọi là **phương sai thực nghiệm** (hay **phương sai mẫu**), là một ước lượng của phương sai của X.

Ví dụ 5.7. **Trung vị** (median) của một biến ngẫu nhiên X là điểm m sao cho $\mathcal{F}_X(m) = P(X \leq m) = 1/2$. Nếu như ảnh ngược $\mathcal{F}_X^{-1}(1/2)$ không phải là một điểm mà là một đoạn thẳng, thì trung vị được định nghĩa là trung điểm của đoạn thẳng đó. Trung vị của X có thể được ước lượng bằng **trung vị thực nghiệm**, tức là của phân bố xác suất thực nghiệm.

Bài tập 5.1. Suy ra định lý 5.1 từ định lý 3.8.

5.2.2 Hàm ước lượng

Giả sử X là một biến ngẫu nhiên có phân bố Poisson với tham số λ. Khi đó ta biết rằng λ vừa là kỳ vọng, vừa là phương sai của X, và như vậy có ít nhất 2 cách khác nhau để ước lượng λ: thông qua kỳ vọng hoặc phương sai của các phân bố thực nghiệm.

Nói một cách tổng quát, giả sử ta muốn ước lượng một đại lượng θ nào đó. Có thể có nhiều cách khác nhau để ước lượng θ, mỗi cách cho bởi một *hàm ước lượng*. Theo định nghĩa, một **hàm ước lượng** (estimator) của θ chẳng qua là một hàm số n biến Θ nào đó, nhận đầu vào là các mẫu thực nghiệm $(x_1, \ldots, x_n)$ của X, và đầu ra là các giá trị **ước lượng** (estimate) của θ:

$$\hat{\theta} = \Theta(x_1, \ldots, x_n). \tag{5.4}$$

Điều chúng ta muốn có là sai số $\hat{\theta} - \theta$ giữa ước lượng $\hat{\theta}$ và giá trị thật của θ càng nhỏ càng tốt. Hay nói cách khác, ước lượng càng chính xác càng tốt.

Nếu g là một hàm số n biến bất kỳ, và $x_1, \ldots, x_n$ là một mẫu thực nghiệm cỡ n của một biến ngẫu nhiên X, thì $g(X_1, \ldots, X_n)$ được gọi là một **hàm thống kê** của biến ngẫu nhiên X, và giá trị $g(x_1, \ldots, x_n)$ được gọi là một **thống kê** (statistic). Như vậy, ta có thể nói rằng, hàm ước lượng (estimator) là một hàm thống kê dùng để ước lượng một đại lượng nào đó, và đại lượng đó được ước lượng bằng thống kê.

Khi cỡ của mẫu thực nghiệm có thể thay đổi, thì ta cần không phải là một, mà là một dãy hàm ước lượng: mỗi hàm cho một cỡ mẫu n. Ta sẽ ký hiệu chung một dãy hàm ước lượng như vậy (để cùng ước lượng một đại lượng θ) bằng một chữ cái (ví dụ Θ) và gọi chung chúng là một hàm. Ta muốn rằng, khi n càng lớn thì nói chung sai số $\hat{\theta} - \theta$ giữa ước lượng $\hat{\theta}$ và giá trị thật của θ phải càng nhỏ. Tính chất này có thể phát biểu chính xác một cách toán học như sau, và gọi là *tính nhất quán*[5] (consistency), của hàm ước lượng:

Định nghĩa 5.1. *Hàm ước lượng Θ của đại lượng θ được gọi là* **nhất quán** *(consistent), nếu như với mọi $\epsilon > 0$ ta có*

$$\lim_{n\to\infty} P(|\Theta(X_1, \ldots, X_n) - \theta| < \epsilon) = 1. \tag{5.5}$$

Tính nhất quán là tính chất quan trọng nhất của hàm ước lượng. Ngoài ra, tùy từng trường hợp, ta có thể đòi hỏi một số tính chất

[5]Có tài liệu gọi tính chất này là *tính vững*, nhưng ở đây chúng ta sẽ dùng từ *nhất quán*, vì từ *vững* tiếng Việt còn được dùng để chỉ một tính chất khác của ước lượng, mà tiếng Anh gọi là *robust*. Các hàm ước lượng vững (robust) là cải tiến của các hàm ước lượng "cổ điển" thường dùng, và cho ước lượng tốt kể cả khi mẫu thực nghiệm chẳng may có những giá trị quá đặc biệt (quá lớn hay quá nhỏ so với thông thường, quá hiếm xảy ra).

khác, ví dụ như tính *không chệch*, hoặc dạng yếu hơn của nó, là tính *không chệch tiệm cận*:

Định nghĩa 5.2. *Hàm ước lượng Θ được gọi là* **không chệch** *(unbiased) nếu như kỳ vọng của $\Theta(X_1, \ldots, X_n)$ bằng θ:*

$$\theta = \mathbb{E}(\Theta(X_1, \ldots, X_n)). \tag{5.6}$$

Hàm ước lượng Θ được gọi là **không chệch tiệm cận** *(asymptotically unbiased) nếu như*

$$\theta = \lim_{n \to \infty} \mathbb{E}(\Theta(X_1, \ldots, X_n)). \tag{5.7}$$

Ví dụ 5.8. Như ta đã thấy trong mục trước, hàm kỳ vọng thực nghiệm

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{5.8}$$

là một hàm ước lượng của kỳ vọng của X. Dễ thấy rằng đây là một hàm ước lượng không chệch, và dạng yếu của luật số lớn nói rằng hàm ước lượng này nhất quán. Moment bậc k của phân bố thực nghiệm cho hàm ước lượng $\frac{1}{n} \sum_{i=1}^{n} X_i^k$ của moment bậc k của X. Hàm ước lượng này nói chung không thỏa mãn tính chất *không chệch* khi $k \geq 2$, nhưng thỏa mãn tính chất không chệch tiệm cận.

Định lý 5.2. *Giả sử Θ là một hàm ước lượng không chệch tiệm cận thỏa mãn điều kiện phương sai tiến đến 0 khi n tiến đến vô cùng:*

$$\lim_{n \to \infty} var(\Theta(X_1, \ldots, X_n)) = 0. \tag{5.9}$$

Khi đó Θ là một hàm ước lượng nhất quán.

Chứng minh. Tương tự như chứng minh của dạng yếu của luật số lớn, suy ra từ bất đẳng thức Chebyschev. □

Ghi chú 5.2. Trong tiếng Việt, nhiều khi thay vì nói “hàm ước lượng” người ta nói đơn giản hóa là “ước lượng” nhưng hiểu là hàm ước lượng. Trong tiếng Anh thì hai từ này không lẫn với nhau: hàm ước lượng gọi là estimator, còn ước lượng gọi là estimate.

Bài tập 5.2. Giả sử X có phân bố đều trên đoạn thẳng $]0,\theta[$. Chứng minh rằng

$$\Theta = \frac{n+1}{n}\max(X_1, X_2, \ldots, X_n) \tag{5.10}$$

là một hàm ước lượng nhất quán không chệch của θ.

Bài tập 5.3. Chứng minh rằng trung vị thực nghiệm là ước lượng nhất quán không chệch tiệm cận của trung vị. Xây dựng ví dụ cho thấy trung vị thực nghiệm nói chung không thỏa mãn tính chất không chệch.

5.2.3 Ước lượng không chệch của phương sai

Hàm phương sai mẫu $\hat{\Sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{\sum_{i=1}^n X_i}{n}\right)^2$ là một ước lượng nhất quán, nhưng *có chệch*, có nghĩa là kỳ vọng của

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{\sum_{i=1}^n X_i}{n}\right)^2$$

không bằng phương sai σ^2 của X. Thật vậy, khi $n = 2$, ta có

$$\mathbb{E}(\hat{\Sigma}^2) = \mathbb{E}\left(\frac{X_1^2 + X_2^2}{2} - \frac{(X_1 + X_2)^2}{4}\right) = \frac{1}{4}\mathbb{E}(X_1^2 + X_2^2 - 2X_1X_2)$$

$$= \frac{1}{4}(\mathbb{E}(X_1^2) + \mathbb{E}(X_2^2) - 2\mathbb{E}(X_1)\mathbb{E}(X_2)) = \frac{1}{4}(2\mathbb{E}(X^2) - 2\mathbb{E}(X)^2) = \frac{1}{2}\sigma^2$$

chứ không bằng σ^2. Tương tự như vậy, khi n tùy ý, có thể kiểm tra rằng $\mathbb{E}(\hat{\Sigma}^2) = \frac{n-1}{n}\sigma^2$. Bởi vậy ta có định nghĩa và định lý sau:

Định nghĩa 5.3. *Hàm*

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \frac{\sum_{i=1}^{n} X_i}{n}\right)^2. \tag{5.11}$$

gọi là **hàm phương sai mẫu hiệu chỉnh**. *Nếu $x_1, \ldots, x_n$ là một mẫu thực nghiệm của X, thì giá trị của S^2 tại bộ điểm $(x_1, \ldots, x_n)$, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)^2$, gọi là* **phương sai mẫu hiệu chỉnh** *(của mẫu $x_1, \ldots, x_n$) của X.*

Định lý 5.3. *Hàm phương sai mẫu hiệu chỉnh là ước lượng không chệch của phương sai σ^2 của biến ngẫu nhiên X.*

Định lý trên giải thích vì sao người ta hay dùng công thức phương sai mẫu hiệu chỉnh $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)^2$, thay vì công thức $\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)^2$, khi nói về phương sai của một bộ n số $(x_1, \ldots, x_n)$. Tỷ lệ giữa phương sai mẫu hiệu chỉnh và phương sai mẫu là $\frac{n}{n-1}$, tiến tới 1 khi n tiến tới vô cùng.

5.2.4 Phương pháp hợp lý cực đại

Phân bố thực nghiệm là một ước lượng của phân bố của X. Nhưng phân bố thực nghiệm luôn luôn là phân bố rời rạc, và có thể không thỏa mãn một số tính chất mà X phải thỏa mãn, tức là không nằm trong họ các phân bố mà X rơi vào, ví dụ như họ các phân bố normal,

họ các phân bố hình học, v.v. Một trong những phương pháp phổ biến nhất để ước lượng phân bố xác suất của X bằng một phân bố xác suất trong một họ nào đó là **phương pháp hợp lý cực đại** (maximal likelyhood - dễ xảy ra nhất).Ý tưởng của phương pháp này là: những gì mà thấy được trong thực nghiệm, thì phải dễ xảy ra hơn là những gì không thấy. Ví dụ như, khi một giáo viên hỏi một học sinh 4 câu hỏi ngẫu nhiên về một môn học nào đó mà học sinh đều trả lời được, thì giáo viên sẽ "ước lượng" rằng đấy là học sinh giỏi, vì khi giỏi thì mới nhiều khả năng trả lời được cả 4 câu hỏi, còn nếu không giỏi sẽ có nhiều khả năng không trả lời được ít nhất 1 trong 4 câu hơn là khả năng "ăn may" trả lời được cả 4 câu. Chúng ta sẽ tìm phân bố xác suất của biến ngẫu nhiên X sao cho mẫu thực nghiệm $(x_1, \ldots, x_n)$ có nhiều khả năng xảy ra nhất.

Ta sẽ giả sử X có phân bố xác suất $P_{\boldsymbol{\theta}}$ phụ thuộc vào một số tham số $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ nào đó. Trong trường hợp $P_{\boldsymbol{\theta}}$ là phân bố rời rạc, ta đặt

$$\mathcal{L}(\boldsymbol{\theta}) = L_{\boldsymbol{\theta}}(x_1, \ldots, x_n) = P_{\boldsymbol{\theta}}(x_1) \ldots P_{\boldsymbol{\theta}}(x_n), \tag{5.12}$$

còn trong trường hợp $P_{\boldsymbol{\theta}}$ là phân bố liên tục với hàm mật độ $\rho_{\boldsymbol{\theta}}$, thì ta đặt

$$\mathcal{L}(\boldsymbol{\theta}) = L_{\boldsymbol{\theta}}(x_1, \ldots, x_n) = \rho_{\boldsymbol{\theta}}(x_1) \ldots \rho_{\boldsymbol{\theta}}(x_n). \tag{5.13}$$

$\mathcal{L}(\boldsymbol{\theta})$ được gọi là **hàm độ hợp lý** (likelyhood function) của $\boldsymbol{\theta}$ (khi mà mẫu $x_1, \ldots, x_n$ đã biết). Bài toán mà chúng ta cần giải, là tìm $\boldsymbol{\theta}$ có độ hợp lý cao nhất, tức là tìm $\hat{\boldsymbol{\theta}}$ sao cho

$$\mathcal{L}(\hat{\boldsymbol{\theta}}) = \sup_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}). \tag{5.14}$$

Với nguyên tắc "đạo hàm bằng 0 tại điểm cực trị", vấn đề tìm điểm cực đại của $\mathcal{L}(\boldsymbol{\theta})$ nhiều khi được đưa về vấn đề giải phương trình:

$$\frac{d}{d\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}) = 0. \tag{5.15}$$

Không phải lúc nào phương pháp hợp lý cực đại cũng cho kết quả, bởi vì chẳng hạn nếu hàm $\mathcal{L}(\boldsymbol{\theta})$ có nhiều điểm cực đại, thì không biết nên chọn điểm nào. Tuy nhiên, trong nhiều bài toán, phương pháp này cho kết quả duy nhất và khá "hợp lý" về trực giác.



Hình 5.4: Ronald Fisher

Ghi chú 5.3. Người khởi xướng phương pháp hợp lý cực đại là Ronald Fisher (1890-1962), một nhà di truyền học và thống kê học người Anh, vào đầu thế kỷ 20. Fisher cùng với Pearson được coi là những cha tổ của thống kê toán học. Khi Fisher đưa ra phương pháp hợp lý cực đại thì Pearson không ủng hộ nó, dẫn đến quan hệ căng thẳng giữa hai người.

Ví dụ 5.9. Giả sử ta biết rằng X phải có phân bố xác suất đều trên một đoạn thẳng $[a, b]$, nhưng ta không biết a và b. Vấn đề đặt ra là ước lượng a và b, dựa trên một mẫu thực nghiệm $x_1, \ldots, x_n$. Ta có

$$\mathcal{L}(a, b) = \rho_{a,b}(x_1) \ldots \rho_{a,b}(x_n) = \frac{1}{(b-a)^n}, \tag{5.16}$$

và ta cần tìm a,b sao cho $1/(b-a)^n$ đạt cực đại. Ta biết rằng các điểm $x_1, \ldots, x_n$ phải nằm trong đoạn thẳng $[a, b]$, như vậy ta phải có $b \geq \max x_i$, $a \leq \min x_i$, và $1/(b-a)^n$ đạt cực đại khi mà $b = \max x_i$, $a = \min x_i$. Bởi vậy các ước lượng của a và b là:

$$\hat{a} = \min x_i,\ \hat{b} = \max x_i. \tag{5.17}$$

Ví dụ 5.10. Giả sử ta muốn tìm xác suất của một sự kiện A nào đó (ví dụ như sự kiện: say rượu khi lái xe). Gọi X là *hàm chỉ báo* của A: $X = 0$ nếu A không xảy ra, và $X = 1$ nếu A xảy ra. Khi đó X có phân bố Bernoulli với tham số $p = P(A)$. Để ước lượng p, ta làm n phép thử ngẫu nhiên độc lập, và được một mẫu $x_1, \ldots, x_n$ của X. Các số $x_1, \ldots, x_n$ chỉ nhận hai giá trị 0 và 1. Gọi k là số số 1 trong dãy số $x_1, \ldots, x_n$, và $n - k$ là số số 0. Khi đó hàm độ hợp lý là:

$$\mathcal{L}(p) = p^k(1-p)^{n-k}. \tag{5.18}$$

Đạo hàm của $\mathcal{L}(p)$ là $\mathcal{L}'(p) = n(k/n - p)p^{k-1}(1-p)^{n-k-1}$, từ đó suy rằng hàm $\mathcal{L}(p)$ đạt cực đại trên đoạn $[0,1]$ tại điểm $p = k/n = \sum_{i=1}^n x_i/n$. Như vậy, theo phương pháp hợp lý cực đại, ta có ước lượng sau đây của xác suất $p = p(A)$:

$$\hat{p} = \sum_{i=1}^{n} x_i/n. \tag{5.19}$$

Ví dụ 5.11. Trở lại bài toán ước lượng phân bố của chiều dài dưa chuột. Ta giả sử X ở đây có phân bố normal $\mathcal{N}(\mu, \sigma^2)$, và ta muốn ước lượng kỳ vọng μ và phương sai σ^2 của X. Theo phương pháp hợp lý cực đại, ta xác định hàm độ hợp lý, khi có một mẫu $x_1, \ldots, x_n$ của X, là:

$$\begin{aligned}\mathcal{L}(\mu,\sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}\right).\end{aligned} \tag{5.20}$$

Để tìm điểm có độ hợp lý cực đại, ta giải hệ phương trình

$$\frac{d}{d\mu}\mathcal{L}(\mu,\sigma) = 0 \quad \text{và} \quad \frac{d}{d\sigma}\mathcal{L}(\mu,\sigma) = 0.$$

Phương trình thứ nhất tương đương với $\frac{d}{d\mu}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$, và cho nghiệm là $\mu = \frac{\sum_{i=1}^n x_i}{n}$. Phương trình thứ hai tương đương với, $\frac{-n}{\sigma} + \frac{2\sum_i (x_i-\mu)^2}{2\sigma^3} = 0$, và cho nghiệm là $\sigma^2 = \frac{\sum_i (x_i-\mu)^2}{n}$. Như vậy, phương pháp hợp lý cực đại cho ta các ước lượng sau đây của kỳ vọng μ và phương sai σ^2:

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n}, \tag{5.21}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)^2. \tag{5.22}$$

Một điều thú vị là đối với phân bố normal, phương pháp hợp lý cực đại cho ta ước lượng của phương sai bằng phương sai mẫu, chứ không bằng phương sai mẫu hiệu chỉnh. (Nhắc lại rằng tỷ lệ giữa hai đại lượng này là $n/(n-1)$).

Bài tập 5.4. Chứng minh rằng các ước lượng trong ví dụ 5.9 thỏa mãn tính chất nhất quán.

Bài tập 5.5. Tìm hàm ước lượng hợp lý cực đại cho tham số λ của phân bố mũ.

Bài tập 5.6. Tìm hàm ước lượng hợp lý cực đại cho tham số λ của phân bố Poisson. (Chú ý: ước lượng này không tồn tại nếu như tất cả các giá trị trong mẫu thực nghiệm đều bằng 0).

5.2.5 Phương pháp moment

Một trong những phương pháp khác hay được dùng để ước lượng phân bố của X bằng một phân bố $P_{\boldsymbol{\theta}}$ nào đó, là giải hệ phương trình sau để tìm ước lượng của các tham số $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$:

$$\int_{\mathbb{R}} x^s dP_{\boldsymbol{\theta}} = \frac{1}{n}\sum_{i=1}^{n} x_i^s \tag{5.23}$$

với mọi $s = 1, \ldots, k$, trong đó $(x_1, \ldots, x_n)$ là một mẫu thực nghiệm của X. Vế bên trái của phương trình trên là moment bậc s của phân bố $P_{\boldsymbol{\theta}}$, còn vế bên phải là moment bậc s thực nghiệm.

Bài tập 5.7. Giả sử X là một biến ngẫu nhiên với phân bố liên tục cho bởi hàm mật độ ρ sau:

$$\rho(x;\lambda) = \begin{cases} \lambda x^{\lambda-1} \text{ nếu } 0 < x < 1 \\ 0 \text{ tại các điểm còn lại} \end{cases}, \tag{5.24}$$

trong đó λ là một tham số. a) Tìm hàm ước lượng hợp lý cực đại của λ.
b) Tìm hàm ước lượng của λ theo phương pháp moment.

5.3 Sai số và độ tin cậy của ước lượng

5.3.1 Sai số của ước lượng

Về nguyên tắc, nói chung mọi ước lượng đều có sai số, bởi vì giá trị của ước lượng phụ thuộc vào hàm hước lượng và giá trị của mẫu thực nghiệm, mà các mẫu thực nghiệm khác nhau của cùng một biến ngẫu nhiên có các giá trị khác nhau, dẫn đến các giá trị ước lượng khác nhau, không thể tất cả đều chính xác được.

Giả sử $\Theta(X_1, \ldots, X_n)$ là một hàm ước lượng của một đại lượng θ nào đó. Trong trường hợp Θ là ước lượng không chệch, tức là kỳ vọng của $\Theta(X_1, \ldots, X_n)$ chính bằng θ, thì ta có thể lấy độ lệch chuẩn của $\Theta(X_1, \ldots, X_n)$ làm thước đo đánh giá mức độ sai số trung bình của một ước lượng của θ dùng hàm ước lượng Θ. Trong trường hợp chung, đại lượng

$$MSE(\Theta) = \mathbb{E}(|\Theta(X_1, \ldots, X_n) - \theta|^2) \tag{5.25}$$

được gọi là **sai số trung bình bình phương** (mean squared error) của hàm ước lượng Θ (cho đại lượng đặc trưng θ của biến ngẫu nhiên X).

Bất đẳng thức Cramér–Rao dưới đây cho ta chặn dưới của sai số trung bình bình phương của các hàm ước lượng. Nó cho thấy, về mặt lý thuyết, khi cỡ của mẫu thực nghiệm là cố định, không thể có cách ước lượng với độ chính xác tùy ý, mà cách ước lượng (không chệch) nào cũng có sai số trung bình bình phương lớn hơn một hằng số nào đó.

Hàm ước lượng có sai số trung bình bình phương càng nhỏ thì được coi là càng hiệu quả (càng chính xác). Hàm ước lượng có sai số trung bình bình phương nhỏ nhất (trong các hàm ước lượng n biến của θ) được gọi là **hàm ước ượng hiệu quả**.

Định nghĩa 5.4. *Giả sử phân bố xác suất $P_X = P_\theta$ nằm trong một họ các phân bố xác suất P_θ phụ thuộc vào tham số θ. Khi đó đại lượng*

$$I(\theta) = \mathbb{E}\left(\left[\frac{\partial \ln L(X,\theta)}{\partial \theta}\right]^2\right) = \int_{\mathbb{R}} \left[\frac{\partial \ln L(x,\theta)}{\partial \theta}\right]^2 dP_\theta, \tag{5.26}$$

trong đó $L(x,\theta) = P_\theta(x)$ trong trường hợp P_θ là phân bố xác xuất rời rạc và $L(x,\theta) = \rho_\theta(x)$ trong trường hợp P_θ là phân bố liên tục với hàm mật độ ρ_θ, được gọi là **lượng thông tin Fisher** *ứng với θ.*

Định lý 5.4 (Bất đẳng thức Cramér–Rao). *Với mọi hàm ước lượng không chệch Θ của θ ta có*

$$\mathbb{E}(|\Theta(X_1,\ldots,X_n) - \theta|^2) \geq \frac{1}{nI(\theta)}. \tag{5.27}$$

Nếu như Θ là ước lượng có chệch, với độ chệch là

$$b(\theta) = \mathbb{E}(\Theta(X_1,\ldots,X_n)) - \theta, \tag{5.28}$$

và ký hiệu $b'(\theta)$ là đạo hàm của $b(\theta)$ theo θ, thì

$$\mathbb{E}(|\Theta(X_1,\ldots,X_n)-\theta|^2) \geq \frac{(1+b'(\theta))^2}{nI(\theta)}. \quad (5.29)$$

Ghi chú 5.4. Trong phát biểu chính xác hơn của định lý trên, cần phải giả sử rằng phân bố xác suất của X thỏa mãn một số điều kiện "regularity" (không kỳ dị) (xem chứng minh phía dưới, sẽ xuất hiện cụ thể điều kiện). Trong các bài toán thực tế, nói chung các điều kiện regularity này luôn được thỏa mãn.

Chứng minh. Ta sẽ chứng minh cho trường hợp $n = 1$, ước lượng là không chệch, và phân bố xác suất là liên tục tuyệt đối với hàm mật độ $\rho_\theta(x) = \rho(\theta, x)$. Trường hợp tổng quát phức tạp hơn, nhưng các chứng minh hoàn toàn tương tự.

Xuất phát từ đẳng thức $\int_{-\infty}^{\infty} \Theta(X)\rho(\theta,x)dx = \theta$ (ước lượng không chệch), lấy đạo hàm theo θ, ta có

$$\int_{-\infty}^{\infty} \Theta(X)\frac{\partial\rho(\theta,x)}{\partial\theta}dx = 1.$$

Chúng ta cần điều kiện không kỳ dị sau: $\displaystyle\int_{-\infty}^{\infty} \frac{\partial\rho(x,\theta)}{\partial\theta}dx = 0$ (ngoài việc tích phân giao hoán với đạo hàm theo θ phía trên). Với điều kiện này, ta có $\displaystyle\int_{-\infty}^{\infty}(\Theta(X)-\theta)\frac{\partial\rho(\theta,x)}{\partial\theta}dx = 1$, hay còn có thể viết

$$\int_{-\infty}^{\infty}(\Theta(X)-\theta)\sqrt{\rho(\theta,x)}\frac{\partial\ln\rho(\theta,x)}{\partial\theta}\sqrt{\rho(\theta,x)}dx = 1.$$

Đẳng thức trên, cùng với bất đẳng thức Cauchy-Schwartz

$$\left(\int_{-\infty}^{\infty} fg dx\right)^2 \leq \left(\int_{-\infty}^{\infty} f^2 dx\right)\left(\int_{-\infty}^{\infty} g^2 dx\right),$$

suy ra
$1 \leq (\int_{-\infty}^{\infty} (\Theta(X) - \theta)\sqrt{\rho(\theta, x)})^2 dx)(\int_{-\infty}^{\infty} (\frac{\partial \ln \rho(\theta, x)}{\partial \theta} \sqrt{\rho(\theta, x)})^2 dx) =$
$\mathbb{E}(|\Theta(X) - \theta|^2).I(\theta)$, và ta được điều phải chứng minh. □

Ghi chú 5.5. Harald Cramér (1893–1985) là nhà toán học và thống kê học Thụy Điển, học trò của nhà toán học Marcel Riesz. Calyampudi Radhakrishna Rao (sinh năm 1920) là nhà thống kê học người gốc Ấn Độ, làm việc tại Mỹ cho đến khi về hưu, học trò của Ronald Fisher. Bất đẳng thức Kramér–Rao được hai ông làm ra vào quãng năm 1945.

Bài tập 5.8. Thử tự chứng minh định lý trên khi n là số tuỳ ý (và X có phân bố liên tục tuyệt đối).

5.3.2 Khoảng tin cậy và độ tin cậy

Vì nói chung mọi ước lượng đều có sai số, nên sau khi tìm được một giá trị ước lượng $\hat{\theta} = \Theta(x_1, \ldots, x_n)$ của θ, ta phải "cho phép" nó có thể có một sai số đến ϵ nào đó, và coi rằng giá trị thật của θ nằm trong đoạn $[\hat{\theta} - \epsilon, \hat{\theta} + \epsilon]$. Đoạn đó gọi là khoảng tin cậy. Nhưng điều đó không có nghĩa là ta tin tưởng 100% rằng θ nằm trong đoạn $[\hat{\theta} - \epsilon, \hat{\theta} + \epsilon]$, mà chỉ có nghĩa là ta tin rằng, với độ tin cậy cao, θ nằm trong khoảng tin cậy $[\hat{\theta} - \epsilon, \hat{\theta} + \epsilon]$. Nói cách khác, ta có

$$P(\theta \in [\hat{\theta} - \epsilon, \hat{\theta} + \epsilon]) = 1 - p, \tag{5.30}$$

trong đó $1 - p$ là **độ tin cậy** (confidence), và $[\hat{\theta} - \epsilon, \hat{\theta} + \epsilon]$ là **khoảng tin cậy** (confidence interval). Tất nhiên, khi khoảng tin cậy càng hẹp (ϵ càng nhỏ), thì độ tin cậy càng thấp Muốn có độ tin cậy cao (tức là p nhỏ), thì cần phải để khoảng tin cậy đủ rộng (ϵ đủ lớn). Với giả sử

ước lượng nhất quán, khi sai số ϵ cố định thì độ tin cậy $1-p$ tiến tới 1 khi cỡ thực nghiệm n tiến tới vô cùng, và ngược lại khi p cố dịnh thì sai số ϵ tiến tới 0 khi n tiến tới vô cùng. Người ta thường hay cố định p (chẳng hạn $p = 5\%$ hay $p = 1\%$), rồi tìm khoảng tin cậy tương ứng cho độ tin cậy đã cố định đó.

Ví dụ 5.12. Giả sử khi đo chiều dài của 100 quả dưa chuột được chọn một cách ngẫu nhiên từ một quần thể (population) các quả dưa chuột sẽ được đóng hộp, ta được các con số sau: $\overline{X}$ (hàm kỳ vọng thực nghiệm) có giá trị bằng 9.3cm (đây là giống dưa chuột nhỏ), và $\hat{\sigma}$ (độ lệch chuẩn thực nghiệm) là 0.5cm. Ta có thể coi là $Z = \dfrac{(\overline{X}-\mu)}{\hat{\sigma}/\sqrt{100}} = \dfrac{(\overline{X}-\mu)}{0.05}$ có phân bố normal chuẩn tắc, trong đó μ là kỳ vọng độ dài của các quả dưa chuột tính theo cm. Đặt $p = 1\%$, ta được $|Z| \leq 2.57$, có nghĩa là $P(|Z| \leq 2.57) \approx 99\% = 1-1\%$. Bất đẳng thức $|\dfrac{(9.3-\mu)}{0.05}| \leq 2.57$ tương đương với $\mu \in [9.3 - 2.57 \times 0.05, 9.3 + 2.57 \times 0.05] \approx [9.17, 9.43]$. Như vậy, $[9.17, 9.43]$ là khoảng tin cậy của μ với độ tin cậy 99%.

Ghi chú 5.6. Có một vấn đề tế nhị trong lý luận trong ví dụ trên, liên quan đến xác suất có điều kiện. Sai số mà chúng ta tính được là $2.57 \times 0.05 \approx 0.13$ với độ tin cậy 99%, có nghĩa là

$$P(|\overline{X}-\mu| < 0.13 \mid \mu \text{ cố định}) \approx 99\%,$$

nhưng sau đó ta lại muốn hiểu điều này thành: $\overline{X} = 9.3$ là cái biết được sau thực nghiệm, μ là cái chưa biết, có thể coi như một biến ngẫu nhiên, và

$$P(|\overline{X}-\mu| < 0.13 \mid \overline{X} = 9.3) \approx 99\%.$$

Vì sao ta có thể coi $P(|\overline{X} - \mu| < 0.13 \mid \mu \text{ cố định}) = P(|\overline{X} - \mu| < 0.13 \mid \overline{X} \text{ cố định})$? Triết lý ở đây là, ta coi rằng phân bố của sai số $\overline{X} - \mu$ không phụ thuộc vào bản thân giá trị của μ, mà chỉ phụ thuộc vào hiệu $\overline{X} - \mu$ (trong lớp những vấn đề đang được xét). Kiểu như khi bắn cung tên vào đích: đặt đích ở đâu không quan trọng, vẫn sẽ có cùng 1 phân bố về độ lệch của mũi tên được bắn so với tâm điểm (μ) của đích. Tất nhiên điều này không hoàn toàn đúng, nhưng đủ gần đúng để ta sử dụng nó. Tương tự như vậy, ta cũng coi rằng phân bố của $\overline{X} - \mu$ không phụ thuộc vào bản thân giá trị của $\overline{X}$, mà chỉ phụ thuộc vào hiệu $\overline{X} - \mu$. Khi đó (trong một không gian xác suất thích ứng cho vấn đề đang được xét) ta có $P(|\overline{X} - \mu| < 0.13 \mid \mu \text{ cố định}) = P(|\overline{X} - \mu| < 0.13) = P(|\overline{X} - \mu| < 0.13 \mid \overline{X} \text{ cố định})$.

Một cách tổng quát hơn, ta có thể thay $\Theta(\mathbf{X}) - \epsilon$ và $\Theta(\mathbf{X}) + \epsilon$ bằng hai thống kê $A = g_1(\mathbf{X})$ và $B = g_2(\mathbf{X})$ bất kỳ ($\mathbf{X} = (X_1, \ldots, X_n)$ là hàm mẫu thực nghiệm của X, và phân bố của X phụ thuộc tham số θ), với $A < B$. Khi đó ta có định nghĩa sau:

Định nghĩa 5.5. *Giả sử $A = g_1(\mathbf{X})$ và $B = g_2(\mathbf{X})$ là hai hàm thống kê, với $A < B$. Giả sử $P(A < \theta < B) = 1 - p$. Khi đó, với mọi giá trị thực nghiệm a của A và b của B (của cùng một mẫu thực nghiệm), ta nói rằng đoạn $]a, b[$ là* **khoảng tin cậy** *của θ với* **độ tin cậy** *$1 - p$, hay còn gọi là* **khoảng tin cậy** *$100(1 - p)\%$* **của** *θ.*

Trong nhiều vấn đề, thay vì ước lượng θ, người ta chỉ muốn đánh giá θ một phía (xem nó lớn hơn, hay nhỏ hơn, cái gì đó). Khi đó người ta dùng các **khoảng tin cậy một phía** $]-\infty, b[$ hay $]a, \infty[$.

Ví dụ 5.13. (*Bầu cử*). Giả sử một cuộc thăm dò ý kiến cho thấy 52% số

người được hỏi, trong số 400 người được chọn một cách ngẫu nhiên trong dân chúng, sẽ bầu cho ứng cử viên tổng thống A trong số 2 ứng cử viên chính. Hỏi có thể nói rằng A sẽ thắng cử với độ tin cậy bằng bao nhiêu? Gọi p là tỷ lệ tổng số người sẽ bầu cho A. Khi đó p cũng là là xác suất để 1 ứng cử viên ngẫu nhiên bầu cho A. Khoảng tin cậy ở đây là một chiều: $p > 50\%$ thì A được bầu. Đại lượng thực nghiệm là $\hat{p} = 52\% = 0.52$, và cỡ thực nghiệm là $n = 400$. Phân bố xác suất ở đây là phân bố Bernoulli, với độ lệch chuẩn là $\sigma^2 = \sqrt{p(1-p)}$. Theo định lý giới hạn trung tâm, ta có

$$P(\frac{\overline{X} - p}{\sigma/\sqrt{n}} < c) \approx \mathcal{F}_{\mathcal{N}(0,1)}(c) = P_{\mathcal{N}(0,1)}(] - \infty, c[).$$

với mọi c. Ta sẽ thay σ bằng độ lệch chuẩn thực nghiệm $\hat{\sigma} = \sqrt{\hat{p}(1-\hat{p})}$, và $\overline{X}$ bằng giá trị $\hat{p} = 0.52$ của nó, trong công thức trên. Như vậy, $\mathcal{F}_{\mathcal{N}(0,1)}(c)$ là độ tin cậy cho khoảng tin cậy một phía

$$]\hat{p} - c\sqrt{\hat{p}(1-\hat{p})/n}, \infty[=]0.52 - 0.02498c, \infty[$$

của p. Để xét khả năng thắng cử, cần xét khoảng tin cậy $]0.5, \infty[$, tức là đặt $0.52 - 0.02498c = 0.5$. Giải phương trình đó, ta được $c = 0.80$, và độ tin cậy là $\mathcal{F}_{\mathcal{N}(0,1)}(0.80) \approx 0.788$. Có nghĩa là, ta có thể dự đoán ứng cử viên A sẽ thắng cử, với độ tin tưởng là 78.8%.

5.3.3 Khoảng tin cậy cho độ lệch chuẩn

Trong các ví dụ ở mục trên, chúng ta đã dùng độ lệch chuẩn thực nghiệm thay thế cho độ lệch chuẩn, khi tính khoảng tin cậy và độ tin cậy cho kỳ vọng. Câu hỏi đặt ra là: việc dùng độ lệch chuẩn thực

nghiệm thay thế cho độ lệch chuẩn có làm giảm sự chính xác của các tính toán nhiều không? Bản thân việc dùng độ lệch chuẩn thực nghiệm làm ước lượng cho độ lệch chuẩni có độ tin cậy và khoảng tin cậy ra sao? Để trả lời câu hỏi đó, ta có thể dùng định lý sau, trong trường hợp phân bố là normal:

Định lý 5.5. *Giả sử $X_1, \ldots, X_n$ là một bộ n biến ngẫu nhiên độc lập có cùng phân bố normal $\mathcal{N}(\mu, \sigma^2)$, và $\overline{X} = (\sum_{i=1}^n X_i)/n$, $\hat{\Sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$. Khi đó $n\hat{\Sigma}^2/\sigma^2$ có phân bố χ^2 với $n-1$ bậc tự do.*

Chứng minh của định lý trên có thể suy ra được dễ dàng từ các tính chất của các phân bố normal nhiều chiều (của vector $(X_1, \ldots, X_n)$), và một phép biến đổi tuyến tính vuông góc, tương tự như trong chứng minh định lý Pearson. Cũng có thể chứng minh bằng cách tính hàm đặc trưng hay hàm sinh moment.

Ví dụ 5.14. Giả sử ta bắt được 20 con rồng. Trung bình mỗi con dài 10 mét, và độ lệch chuẩn thực nghiệm của mẫu 20 con đó là 1 mét. Tính khoảng tin cậy 90% của độ lệch chuẩn của chiều dài của rồng? Ta coi chiều dài của rồng có phân bố normal với độ lệch chuẩn σ, và gọi hàm độ lệch chuẩn thực nghiệm của một mẫu 20 con rồng là $\hat{\Sigma}$. Theo định lý trên, $20\hat{\Sigma}^2/\sigma^2$ có phân bố χ^2_{19} với 19 bậc tự do. Để tìm một khoảng tin cậy cho σ với độ tin cậy 90%, ta có thể tìm hai số $\chi^2_{0.05}$ và $\chi^2_{0.95}$ sao cho

$$P(\chi^2_{19} \leq \chi^2_{0.05}) = 5\% \text{ và } P(\chi^2_{19} \geq \chi^2_{0.95}) = 5\% \tag{5.31}$$

Dùng máy tính hoặc tra bảng, ta tìm được $\chi^2_{0.05} \approx 10.12$ và $\chi^2_{0.95} \approx 30.14$, bởi vậy

$$P(10.12 < \frac{20\hat{\Sigma}^2}{\sigma^2} < 30.14) \approx 1 - 5\% - 5\% = 90\%. \tag{5.32}$$

Với giá trị thực nghiệm $\hat{\Sigma} = 1$, ta được bất đẳng thức

$$10.12 < \frac{20}{\sigma^2} < 30.14, \qquad (5.33)$$

tương đương với

$$0.81 < \sigma < 1.41. \qquad (5.34)$$

Như vậy, khoảng tin cậy 90% cho σ là $]0.81, 1.41[$. Có thể thấy đây là một khoảng khá rộng (chênh lệch gần 2 lần giữa số đầu và số cuối). Lý do là vì $n = 20$ tương đối nhỏ, nên độ chính xác của ước lượng độ lệch chuẩn không cao.

5.3.4 Phân bố Student

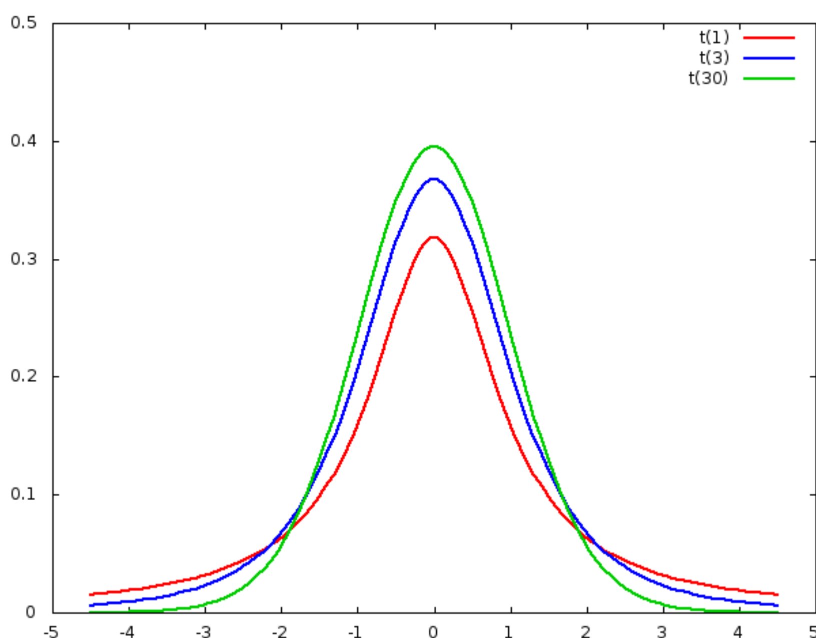
Trong các ví dụ ở mục trên, ta dùng định lý giới hạn trung tâm, rồi thay độ lệch chuẩn bằng độ lệch chuẩn thực nghiệm $\hat{\sigma}$, để kết luận rằng phân bố xác suất của $\frac{\overline{X} - p}{\hat{\sigma}/\sqrt{n}}$ (hay của $\frac{\overline{X} - p}{\hat{\Sigma}/\sqrt{n}}$, trong đó $\hat{\Sigma}^2$ là hàm phương sai thực nghiệm, còn $\hat{\sigma}^2$ là một giá trị thực nghiệm của nó), có thể xấp xỉ bằng phân bố normal chuẩn tắc $\mathcal{N}(0, 1)$. Điều này chỉ đúng đắn khi mà n đủ lớn. Khi n nhỏ thì xấp xỉ này không còn tốt nữa, và khi đó thì thay vì phân bố normal chuẩn tắc ta phải dùng các mô hình phân bố khác. Bởi vậy ta có định nghĩa sau:

Định nghĩa 5.6. *Nếu $X_1, \ldots, X_n$ là một bộ n biến ngẫu nhiên độc lập có cùng phân bố normal $\mathcal{N}(\mu, \sigma^2)$, và*

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}, \qquad (5.35)$$

trong đó $\overline{X} = (\sum_{i=1}^n X_i)/n$ và $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2}$, thì phân bố xác suất của T được gọi là **phân bố Student**, *hay* **phân bố T** *(Student T-distribution), với $n - 1$ bậc tự do.*

Ghi chú 5.7. Dễ thấy rằng, trong định nghĩa trên, phân bố của T không phụ thuộc vào μ và σ. Phân bố T được nhà thống kê học người Anh, ông William Sealy Gosset (1876–1937), đưa ra vào năm 1908, khi đang làm việc cho hãng bia Guinness ở Dublin (thống kê để chọn bia ngon). Do nguyên tắc giữ bí mật của hãng bia, Gosset không được phép ký tên các bài báo của mình với tên thật, nên lấy bút danh là Student. Khái niệm *bậc tự do* của phân bố T là do Ronald Fisher đưa ra, vì nó phù hợp với các công trình khác của Fisher liên quan đến bậc tự do.



Hình 5.5: Hàm mật độ của các phân bố T với 1, 3 và 30 bậc tự do

Phân bố Student rất quan trọng trong việc xác định các khoảng tin cậy và độ tin cậy trong trường hợp mẫu thực nghiệm có cỡ nhỏ. Bởi vậy nó được nghiên cứu khá kỹ lưỡng. Công thức để tính hàm mật độ của phân bố Student là:

Định lý 5.6. *Phân bố Student T với $\nu \geq 1$ bậc tự do có hàm mật độ sau:*

$$\rho_\nu(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}} \frac{1}{(1+\frac{x^2}{\nu})^{\frac{\nu+1}{2}}}. \tag{5.36}$$

Công thức trên có thể suy ra được từ công thức tính hàm mật độ của phân bố χ^2 và từ định lý sau.

Định lý 5.7. *Giả sử $X_1, \ldots, X_n$ là một bộ n biến ngẫu nhiên độc lập có cùng phân bố normal $\mathcal{N}(\mu, \sigma^2)$, và $\overline{X} = (\sum_{i=1}^n X_i)/n$, $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2}$. Khi đó:*
i) $\overline{X}$ độc lập với các biến ngẫu nhiên $X_i - \overline{X}$ và với S.
ii) $(n-1)S^2/\sigma^2$ có phân bố χ^2 với $n-1$ bậc tự do.
iii) Nếu Z có phân bố $\mathcal{N}(0,1)$ và U có phân bố χ^2 với m bậc tự do, thì $Z\sqrt{m}/\sqrt{U}$ có phân bố Student T với m bậc tự do.

Chứng minh của định lý trên có thể suy ra được dễ dàng từ các tính chất của các phân bố normal nhiều chiều.

Theo định lý giới hạn trung tâm, thì phân bố Student với n bậc tự do hội tụ đến phân bố normal chuẩn tắc khi n tiến tới vô cùng. Tuy nhiên, khi n nhỏ, thì độ chênh lệch giữa phân bố Student và phân bố normal chuẩn tắc khá cao. Hình 5.5 là độ thị hàm mật độ của các phân bố t với 1, 3 và 30 bậc tự do. Khi số bậc tự do là 30 trở lên thì phân bố t gần bằng phân bố normal chuẩn tắc, nhưng với số bậc tự

do nhỏ nhỏ, thì nó "dàn trải" hơn nhiều so với phân bố normal chuẩn tắc. Có thể tính toán các giá trị của hàm phân phối xác suất của phân bố Student T bằng cách tra bảng hoặc dùng chương trình máy tính.

Ví dụ 5.15. Tiếp tụ ví dụ trong mục trước về rồng. Vì việc xấp xỉ độ lệch chuẩn bằng độ lệch chuẩn thực nghiệm có độ chính xác kém khi $n = 20$, nên ta phải dùng phân bố T thay cho phân bố normal chuẩn tắc khi tính khoảng tin cậy của kỳ vọng. Đặt $T = (\overline{X} - \mu)\sqrt{n-1}/\hat{\Sigma}$, trong đó $n = 20$ là số con rồng bắt được (cỡ của mẫu), μ là kỳ vọng chiều dài của rồng, $\overline{X}$ là hàm kỳ vọng thực nghiệm, và $\hat{\Sigma}$ là hàm độ lệch chuẩn thực nghiệm. Khi đó T tuân theo phân bố Student T với 19 bậc tự do. Giả sử ta muốn tìm khoảng tin cậy 90% cho μ. Ta cần tìm số c sao cho

$$P(|T| < c) = 90\%. \tag{5.37}$$

Tra theo phân bố Student T với 19 bậc tự do, ta có $c \approx 1.729$. Bởi vậy ta cần giải bất phương trình

$$|(\overline{X} - \mu)\sqrt{20-1}/\hat{\Sigma}| < 1.729, \tag{5.38}$$

trong đó $\overline{X} = 10$ và $\hat{\Sigma} = 1$ (là các đại lượng thực nghiệm). Kết quả là

$$9.603 < \mu < 10.397 \tag{5.39}$$

với độ tin cậy 90%.

Bài tập 5.9. Giả sử có một loại xe ô tô mới, cho 5 người chạy thử 5 xe khác nhau trên đường cao tốc, với kết quả chạy 100km hết lần lượt là 4.53, 3.82, 4.37, 3.91, 4.16 lít xăng. Tìm khoảng tin cậy cho số lít xăng tiêu tốn trung bình của loại xe này cho 100km đường cao tốc, với độ tin cậy 90%.

5.4 Kiểm định các giả thuyết

Trong phần này, và phần sau, chúng ta sẽ bàn đến những phương pháp thống kê dùng để tra lời những câu hỏi dạng "có hay không một hiện tượng hay hiệu ứng nào đó". Ví dụ: loại thuốc chữa bệnh cảm này có hiệu nghiệm không?, có kỳ thị giới tính trong việc tuyển người không?, chất thải của nhà máy này có làm hại sức khỏe của nhân dân xung quanh không?, sở thích âm nhạc có thay đổi theo độ tuổi không?, độc quyền có ảnh hưởng xấu đến kinh tế không?, v.v. Mỗi tình huống "có hay không" như vậy có thể viết dưới hạng một giả thuyết, thường ký hiệu là H_0, gọi là **không thuyết** (null hypothesis), và một giả thuyết đối ngược lại nó, thường ký hiệu là H_1 hoặc H_{alt}, gọi là **đối thuyết** (alternative hypothesis).

Có một điều mà bạn đọc cần hết sức chú ý. Đó là, mỗi phương pháp kiểm định bằng thống kê chỉ thích hợp trong những tình huống nhất định, khi các giả sử nhất định được thoả mãn. Khi có một vấn đề kiểm định thống kê trong thực tế cần thực hiện, thì phải chọn lựa phương pháp đúng đắn, và rất có thể là phương pháp mà bạn đọc cần đến không nằm trong quyển sách này (vì số phương pháp thì nhiều, mà quyển sách chỉ giới thiệu một số phương pháp cơ sở), và bạn đọc sẽ phải tìm hiểu sâu thêm về thống kê để chọn lựa được phương pháp thích hợp cho vấn đề của mình.

5.4.1 Một số nguyên tắc chung của kiểm định bằng thống kê

Tương tự như ước lượng, việc kiểm định giả thuyết bằng thống kê không cho kết quả "chính xác 100%", mà chỉ cho kết quả với một độ tin cậy nhất định nào đó, và có thể xảy ra sai lầm. Các sai lầm có thể phân làm hai loại:
- **Sai lầm loại 1**: phủ nhận giả thuyết H_0, chấp nhận đối thuyết H_1, trong khi H_0 đúng
- **Sai lầm loại 2**: giữ giả thuyết H_0, không chấp nhận đối thuyết H_1, trong khi H_1 đúng.

Cả hai loại sai lầm đều có thể gây ra những hậu quả không tốt. Tùy từng trường hợp mà đánh giá xem sai lầm loại nào dẫn đến hậu quả nghiêm trọng hơn, và cần tránh hơn. Ví dụ, trong trường hợp chất thải có thể gây ung thư: nếu theo thống kê, H_0 xảy ra với độ tin tưởng 80% (tức là với độ tin tưởng 80%, chất thải không gây ung thư) và chỉ có 20% là H_1 (chất thải gây ung thư) xảy ra, thì như thế cũng đủ quá nguy hiểm với tính mạng con người, và trong trường hợp này không chấp nhận được H_0 (tức là không thể để cho nhà máy thải chất thải như vậy). Nhưng ngược lại, nếu đối với một loại thuốc mới, kiểm định thống kê cho thấy H_0 (giả thuyết thuốc không có tác dụng) có trên 5% khả năng xảy ra, thì nói chung thuốc chưa được Bộ Y Tế của các nước chấp nhận, và phải nghiên cứu và thí nghiệm thêm cho đến khi chứng tỏ được là H_1 (giả thuyết thuốc có tác dụng) là đúng đắn với độ tin tưởng rất cao (ít ra trên 95%) thì thuốc mới được chấp nhận.

Khi kiểm định bằng thống kê, các giả thuyết và đối thuyết thường có thể phát biểu lại dưới dạng: một đại lượng nào đó (mà ta không biết, muốn ước lượng) nằm trong một đoạn thẳng nào đó, với độ tin cậy nào đó. Bởi vậy, *các bài toán kiểm định có thể coi như là những trường hợp đặc biệt của các bài toán ước lượng*. Ví dụ, nếu H_0 là, "khi bầu vào quốc hội, đàn bà cũng có xác suất được bầu nhiều như đàn ông", thì H_0 có thể bị loại bỏ và H_1 được chấp nhận nếu như ước lượng cho thấy "xác suất để người được bầu vào quốc hội là đàn ông" nằm trong đoạn $]1/2, \infty[$, với độ tin cậy trên 99%.

Nhắc lại rằng, trong vấn đề ước lượng, độ tin cậy được coi bằng xác suất để một kết quả thống kê thực nghiệm nằm trong một miền nào đó, khi mà đại lượng mà ta muốn ước lượng nằm trong một khoảng nào đó (khoảng tin cậy). Ta đạt độ tin cậy đó, khi mà kết quả thống kê thực nghiệm của ta nằm trong miền cần thiết. Trong ví dụ bầu cứ quốc hội, thì đại lượng mà ta muốn ước lượng là xác suất để một người được bầu cử là đàn bà. Có hai cách phát biểu điều kiện: hoặc là "xác suất để người được bầu là đàn bà không nhỏ hơn 50%" hoặc là "xác suất để người được bầu là đàn bà bằng 50%". Giả sử kết quả thống kê ở đây là số người được bầu vào quốc hội là đàn ông là một số N. Khi đó xác suất (độ tin cậy) ở đây có thể viết là

$$P_1 = P(\text{x.s. để ng. được bầu là nữ} \geq 50\% \mid \text{số ng. được bầu là nam} = N),$$

hoặc là

$$P = P(\text{số ng. được bầu là nam} \geq N \mid \text{x.s. để ng. được bầu là nữ} = 50\%),$$

Đại lượng P cuối cùng là cái mà ta có thể tính được trực tiếp bằng các công thức xác suất, còn P_1 là độ tin cậy, ta không tính trực tiếp, mà lý luận rằng nó có thể coi là (gần) bằng P. Chú ý rằng ta không

viết

$$P(\text{x.s. để ng. được bầu là nữ} = 50\% \mid \text{số ng. được bầu là nam} = N),$$

hoặc là

$$P(\text{số ng. được bầu là nam} = N \mid \text{x.s. để ng. được bầu là nữ} = 50\%),$$

vì nếu dùng đẳng thức ở cả sự kiện và điều kiện, thì xác suất nói chung sẽ rất nhỏ, dù thực tế xảy ra thế nào, và bởi vậy không dùng nó để kiểm định được.

Giá trị $P =$ (xác suất để số người được bầu là đàn ông $\geq N$ dưới điều kiện: xác suất để người được bầu là đàn bà $= 50\%$) được gọi là **giá trị P** cho giả thuyết H_0 (xác suất để người được bầu là đàn bà = 50%). Nó là xác suất sao cho giá trị của thống kê (ở đây là số đàn ông được bầu) bằng hoặc thái quá hơn là giá trị thực nghiệm nhận được (ở đây là số N). Trong trường hợp chung, ta có định nghĩa sau:

Định nghĩa 5.7. Giá trị P *(P-value) là xác suất để giá trị của một thống kê nào đó rơi vào một miền nào đó, khi mà giả thuyết H_0 đúng:*

$$P = P(G \in A \mid H_0), \tag{5.40}$$

trong đó G là một thống kê và A là miền gồm những giá trị bằng hoặc thái quá hơn so với giá trị thực nghiệm của G.

Nguyên tắc kiểm định như sau: *Cố định một số α nào đó (ví dụ $\alpha = 1\%$ hoặc $\alpha = 5\%$). Nếu giá trị P nhỏ hơn α thì chấp nhận đối thuyết H_1, còn nếu $P \geq \alpha$ thì giữ giả thuyết H_0.*

Ví dụ 5.16. Tung một đồng tiền 20 lần, ra 2 lần mặt sấp và 18 lần mặt ngửa. Có thể coi đồng tiền là cân bằng (hai mặt sấp và ngửa

đều có xác suất 50%) không? Giả thuyết H_0 là "đồng tiền cân bằng". Gọi X là biến ngẫu nhiên "số lần hiện mặt sấp trong 20 lần tung". Giá trị P ở đây là: $P = P(X \leq 2 \mid \text{xác suất hiện mặt sấp} = 50\%) = (C_{20}^0 + C_{20}^1 + C_{20}^2)/2^{20} \approx 0.02\%$. Giá trị này quá nhỏ để có thể chấp nhận giả thuyết H_0.

Giá trị P ở phía trên có thể coi là xác suất để xảy ra sai lầm loại 1. Nó thích hợp cho những trường hợp mà sai lầm loại 1 là cái cần chú ý đến (hơn so với sai lầm loại 2). Nếu cần chú ý đến sai lầm loại 2, thì phải tính xác suất để xảy ra sai lầm loại 2, thay vì xác suất để xảy ra sai lầm loại 1. (Phương pháp làm hoàn toàn tương tự).

Bài tập 5.10. (*Tuổi lấy chồng ở Roma thời cổ đại*). Có một lý thuyết của các nhà khảo cổ học cho rằng, tuổi lấy chồng lần đầu trung bình ở Roma thời cổ đại là khoảng 19 tuổi, vì ở các mộ phụ nữ mà có văn bia (epitaph) là do người cha viết thì tuổi trung bình dưới 19, con do chồng viết thì tuổi trung bình trên 19. (Người ta giả thuyết rằng phụ nữ đã có chồng khi chết thì văn bia do chồng viết, còn chưa có chồng thì do cha viết). Thế nhưng, theo một ghi chép lịch sử, tuổi lấy chồng lần đầu của 26 phụ nữ Roma cổ đại được ghi là: 11, 12, 12, 12, 12, 13, 13, 13, 13, 13, 14, 14, 14, 14, 14, 14, 15, 15, 15, 15, 15, 15, 16, 16, 17, 17[(6)]. Gọi H_0 là giả thuyết "tuổi lấy chồng trung bình ở Roma cổ đại là 19", và giả sử rằng tuổi lấy chồng có phân bố normal, với độ lệch chuẩn coi bằng độ lệch chuẩn mẫu của mẫu thực nghiệm với 26 giá trị trên, tức là bằng 1.57. Chứng minh rằng giá trị P nhỏ hơn 1% (Gợi ý: có thể dùng bất đẳng thức Chebyschev).

(6) Theo: A. Lelis, W. Percy, B. Verstraete, *The age of first mariage in ancient Rome*, Edwil Mellen Press, 2003; trích lại từ [7].

5.4.2 Kiểm định Z và kiểm định T cho kỳ vọng

Bởi vì bài toán kiểm định có thể coi là trường hợp đặc biệt của bài toán ước lượng, nên các phân bố hay được dùng để tính khoảng tin cậy trong ước lượng cũng xuất hiện trong kiểm định. Hai loại phân bố hay gặp nhất là: phân bố normal chuẩn tắc (dùng trong trường hợp mẫu thực nghiệm lớn), và phân bố T (cho mẫu thực nghiệm nhỏ, với giả sử là phân bố xác suất ban đầu là normal hặc gần giống normal). Các kiểm định dùng phân bố normal chuẩn tắc được gọi là **kiểm định Z**, còn các kiểm định dùng phân bố Student T được gọi là **kiểm định T**. Chẳng hạn, ta có định nghĩa sau:

Định nghĩa 5.8. Kiểm định Z *cho giá trị kỳ vọng là kiểm định giả thuyết dùng thống kê*

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \text{ hoặc là } Z = \frac{\overline{X} - \mu_0}{\hat{\sigma}/\sqrt{n}}, \tag{5.41}$$

trong đó:
i) $\overline{X}$ là giá trị trung bình của một mẫu thực nghiệm cỡ n của một biến ngẫu nhiên X
ii) Giả thuyết ở đây là về kỳ vọng $\mu = \mathbb{E}(X)$ của X. Giả thuyết H_0 là $\mu = \mu_0$, và đối thuyết là $\mu \neq \mu_0$ (hoặc là $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$; hoặc là $H_0 : \mu \geq \mu_0, H_1 : \mu < \mu_0$).
iii) Hoặc là X được giả sử là có phân bố normal với độ lệch chuẩn σ biết trước, hoặc là n đủ lớn sao cho có thể coi là Z có phân bố normal chuẩn tắc, và thay vì dùng σ ta dùng $\hat{\sigma}$, trong đó $\hat{\sigma}^2$ là phương sai thực nghiệm (cho bởi công thức (5.3)).

Tương tự như vậy, có thể định nghĩa *kiểm định T cho giá trị kỳ*

vọng, tức là kiểm định dùng thống kê $T = \dfrac{\overline{X} - \mu_0}{\hat{\Sigma}/\sqrt{n}}$, và coi rằng T có phân bố Student T với $n - 1$ bậc tự do.

Ví dụ 5.17. (*Thuốc cảm*). Giả sử một hãng dược phẩm muốn kiểm định sự hiệu nghiệm của một loại thuốc cảm mới. Thuốc được đưa cho 100 người bệnh ngẫu nhiên sử dụng khi bắt đầu có triệu chứng cảm. Giả sử một người bị cảm mà không chữa bằng thuốc, thì quá trình bị cảm kéo dài trung bình 7 ngày. Gọi $\overline{X}$ là Giả sử độ dài trung bình của đợt bị cảm của những người được thử cho dùng thuốc mới là $\overline{X} = 5.3$, với độ lệch chuẩn thực nghiệm là 1.5 ngày. Hỏi thông tin này có đủ để chứng tỏ thuốc có hiệu nghiệm không?

Vì $\overline{X} = 5.3 < 7$ nên chúng ta muốn chấp nhận đối thuyết H_1 (thuốc hiệu nghiệm). Nhưng trước khi chấp nhận nó, chúng ta cần phải khẳng định được rằng giá trị P ở đây,

$$P = P(\overline{X} \leq 5.3 | H_0),$$

rất nhỏ. Giả thuyết H_0 ở đây có thể hiểu là sự kiện "$\mathbb{E}(X) = 7$", tức là nếu đem thuốc mới dùng đại trà, thì kỳ vọng độ kéo dài của đợt cảm không khác gì so với nếu không dùng thuốc). Thay vì tính $P(\overline{X} \leq 5.3 | \mathbb{E}(X) = 7)$, ta có thể chỉ cần kiểm tra xem $P(\overline{X} \leq 5.3 | \mathbb{E}(X) = 7) < \alpha$ hay không, trong đó α là một số rất nhỏ nào đó, ví dụ $\alpha = 1\%$. Để làm điều đó, ta cần giải phương trình

$$P(\overline{X} \leq c | \mathbb{E}(X) = 7) = \alpha, \tag{5.42}$$

rồi kiểm tra xem điều kiện $5.3 < c$ có được thỏa mãn không.

Mẫu thực nghiệm ở đây là đủ lớn ($n = 100$) để áp dụng định lý giới hạn trung tâm và dùng *kiểm định Z*. Nói cách khác, ta có thể coi

$Z = \frac{\sqrt{100}(\overline{X} - \mathbb{E}(X))}{\sigma(X)} = \frac{10(\overline{X} - \mathbb{E}(X))}{\sigma(X)}$ là một biến ngẫu nhiên với phân bố normal chuẩn tắc $\mathcal{N}(0,1)$. Khi đó $\overline{X} \leq c$ tương đương với $Z \leq \frac{10(c-\mathbb{E}(X))}{\sigma(X)}$. Ta sẽ coi $\sigma(X)$ bằng độ lệch chuẩn thực nghiệm, tức là $\sigma(X) = 1.5$, và $\mathbb{E}(X) = 7$. Như vậy

$$\alpha \approx P(Z \leq \frac{10(c-7)}{1.5}) \tag{5.43}$$

là giá trị của hàm phân phối xác suất của phân bố normal chuẩn tắc tại điểm $\frac{10(c-7)}{1.5}$. Đặt $\alpha = 1\%$, ta được $\frac{10(c-7)}{1.5} \approx -2.33$, tức là $c \approx (7 - 2.33) \times 1.5/10 \approx 6.65$. Vì $5.3 < 6.65$, nên $P < \alpha = 1^{\%}$, và ta có thể chấp nhận đối thuyết H_1, tức là thuốc có hiệu nghiệm.

Bài tập 5.11. Trong một trang trại nuôi bò lớn, trọng lượng trung bình của bò là 520kg. Một loại thực đơn mới nhằm tăng trọng lượng cho bò được đem thử trên 50 con bò chọn ngẫu nhiên. Các con bò được thử đạt trọng lượng trung bình là 528kg với độ lệch chuẩn 25kg. Hỏi thực đơn mới có hiệu nghiệm không? (Dùng kiểm định Z).

Bài tập 5.12. Một hãng xe ô tô tuyên bố là một loại xe mới do hãng sản xuất chỉ tiêu tốn trung bình 3.0 lít xăng cho 100km trên đường cao tốc. Một tổ chức độc lập kiểm tra khẳng định này, bằng cách cho 5 người chạy thử 5 xe khác nhau của loại xe mới đó, và kết quả là: 2.90, 2.95, 3.10, 3.35, 3.45 (lít/100km). Dựa theo số liệu này, hãy xác định xem tuyên bố của hãng xe ô tô có chấp nhận được không? Giả sử cho 5 người khác chạy thử thêm 5 xe, và được thêm 5 kết quả là 2.95, 3.00, 3.15, 3.30, 3.40. Kiểm định lại xem tuyên bố của hãng xe ô tô có chấp nhận được không, dựa trên tổng cộng 10 kết quả. (Dùng kiểm định T).

Bài tập 5.13. Các trạm cung cấp nước cho thành phố phải kiểm tra chất lượng nước hàng giờ, trong đó có kiểm tra độ pH. Mục tiêu là giữ độ pH của nước ở quãng 8.5 (hơi có tính kiềm: trên 7 là kiềm, dưới 7 là axít). Một lần kiểm tra 15 mẫu nước ở một trạm, thấy rằng độ pH trung bình của các mẫu bằng 8.28 và độ lệch chuẩn là 0.14. Hỏi rằng có đủ cơ sở để kết luận rằng độ pH trung bình của nước ở đó vào thời điểm đó khác 8.5? (Dùng kiểm định T).

5.4.3 Kiểm định so sánh hai kỳ vọng

Giả sử ta muốn so sánh kỳ vọng của hai biến ngẫu nhiên X và Y với nhau, dựa trên một mẫu thực nghiệm cỡ n_X của X và một mẫu thực nghiệm cỡ n_Y của Y. Giả thuyết H_0 là $\mathbb{E}(X) = \mathbb{E}(Y) + \Delta$ (hoặc $\mathbb{E}(X) \leq \mathbb{E}(Y) + \Delta$) và đối thuyết H_1 là $\mathbb{E}(X) \neq \mathbb{E}(Y) + \Delta$ (hoặc $\mathbb{E}(X) > \mathbb{E}(Y) + \Delta$). Ở đây Δ là độ chênh lệch giữa hai kỳ vọng theo giả thuyết.

Khi n và m lớn, thì dựa theo định lý giới hạn trung tâm và ước lượng của độ lệch chuẩn, ta có thể coi rằng $\frac{\overline{X} - \mathbb{E}(X)}{\hat{\Sigma}_X/\sqrt{n_X}}$ và $\frac{\overline{Y} - \mathbb{E}(Y)}{\hat{\Sigma}_Y/\sqrt{n_Y}}$ là hai biến ngẫu nhiên độc lập có phân bố normal chuẩn tắc, trong đó $\overline{X}$ là hàm kỳ vọng thực nghiễm của X (với cỡ thực nghiệm n_X), $\hat{\Sigma}_X$ là hàm độ lệch chuẩn thực nghiệm của X, và tương tự như vậy cho Y. Từ đó suy ra ta cũng có thể coi rằng $\frac{(\overline{X} - \mathbb{E}(X)) - (\overline{Y} - \mathbb{E}(Y))}{\sqrt{\frac{\hat{\Sigma}_X^2}{n_X} + \frac{\hat{\Sigma}_Y^2}{n_Y}}}$ có

phân bố normal chuẩn tắc. Bởi vậy ta có thể đặt

$$Z = \frac{\overline{X} - \overline{Y} - \Delta}{\sqrt{\frac{\hat{\Sigma}_X^2}{n_X} + \frac{\hat{\Sigma}_Y^2}{n_Y}}}. \tag{5.44}$$

Nếu giả thuyết H_0 là đúng, tức là $\mathbb{E}(X) = \mathbb{E}(Y) + \Delta$ thì Z có phân bố normal chuẩn tắc. Kiểm định dựa trên giá trị thực nghiệm của thống kê Z này được gọi là **kiểm định Z hai mẫu** (two sample Z test) để so sánh hai kỳ vọng.

Ví dụ 5.18. (*Aspirin chống đau tim*). Trong một đợt thử nghiệm lớn, 22071 bác sĩ tham gia thử nghiệm lâm sàng về tác dụng của Aspirin chống đau tim. Các bác sĩ được chia một cách ngẫu nhiên thành hai nhóm: nhóm 1 gồm 11037 người, được cho dùng Aspirin, còn nhóm hai gồm 11034 người được cho dùng placebo (không có thuốc). Không ai được biết mình thuộc nhóm được cho thuốc hai là nhóm placebo. Kết quả thử nghiệm cho thấy: 104 người thuộc nhóm dùng aspirin bị lên cơn đau tim (heart attack), và nhóm placebo có 189 người bị lên cơn đau tim, những người còn lại không bị. Hỏi thuốc có hiệu nghiệm để chống đau tim không?

Trong bài toán này, có thể đặt X là biến Bernoulli, bằng 1 nếu bị đau tim, bằng 0 nếu không bị đau tim, trên quần thể những người không dùng aspirin. Kỳ vọng của X là xác suất để bị đau tim với điều kiện là không dùng aspirin. Biến Y tương tự, nhưng cho những người có dùng aspirin. Thuốc hiệu nghiệm nếu kiểm định cho thấy kỳ vọng của Y phải nhỏ hơn kỳ vọng của X. Có thể coi H_0 là giả thuyết kỳ vọng của Y bằng kỳ vọng của X. Ta có một mẫu thực nghiệm của X với cỡ 11034, kỳ vọnng thực nghiệm là 189/11034, và phương sai

mẫu là (1 - 189/11034).189/11034. Tương tự như vậy cho Y. Giá trị của thống kê Z bằng:

$$\frac{\frac{189}{11034} - \frac{104}{11037}}{\sqrt{\frac{(1-189/11034).189/11034}{11034} + \frac{(1-104/11037).104/11037}{11037}}} \approx 5.$$

Vì $P = P_{\mathcal{N}(0,1)}([5, +\infty[) < 1/10^6$ là con số quá nhỏ, nên ta có thể dễ dàng loại bỏ H_0 và chấp nhận đối thuyết H_1, tức là aspirin có hiệu nghiệm chống lên cơn đau tim.

Trong trường hợp mà mẫu thực nghiệm của X và Y có cỡ nhỏ(các số n_X và n_Y nhỏ), kiểm định Z không còn chính xác. Có thể thay thế nó bằng **kiểm định T hai mẫu** (two sample T test), nếu như X và Y có phân bố (gần giống) phân bố normal, và được coi là có *phương sai bằng nhau*. Thống kê T ở đây là

$$T = \frac{\overline{X} - \overline{Y} - \Delta}{\sqrt{\frac{\hat{\Sigma}_X^2}{n_Y} + \frac{\hat{\Sigma}_Y^2}{n_X}} . \sqrt{\frac{n_X + n_Y}{n_X + n_Y - 2}}}. \qquad (5.45)$$

Với giả sử rằng H_0 là đúng ($\mathbb{E}(X) - \mathbb{E}(Y) = \Delta$), thì T có phân bố Student T với $n_X + n_Y - 2$ bậc tự do$^{(7)}$.

Ví dụ 5.19. Giả sử một người nghiên cứu xã hội muốn điều tra xem những người trẻ độ tuổi 20-30 và những người già độ tuổi trên 70, có hài lòng về cuộc sống hiện tại như nhau không. Người này phỏng vấn ngẫu nhiên 10 người già và 12 người trẻ, và đánh giá độ hài lòng theo thang điểm từ 0 đến 100 (100 là hoàn toàn hài lòng). Giả sử các

---

$^{(7)}$Nếu X và Y có phương sai khác nhau, thì có một kiểm định T tương tự, gọi là *Welch's T test*, cũng dùng phân bố T, nhưng với số bậc tự do được tính một cách phức tạp hơn.

kết quả nhận được là:
Người trẻ: 77, 68, 82, 55, 91, 63, 78, 56, 47, 80, 78, 60;
Người già: 76, 35, 66, 53, 85, 38, 47, 66, 72, 61.
Giả thuyết H_0 là người trẻ và người già có độ hài lòng về cuộc sống như nhau. Mẫu thực nghiệm ở đây tương đối nhỏ, không thích hợp cho kiểm định Z, nhưng ta có thể dùng kiểm định T. Gọi X là biến "độ hài lòng của một người trẻ", Y là biến "độ hài lòng của một người già". Theo hai mẫu thực nghiệm trên, ta có:

$$n_X = 12, \overline{X} = 69.58, \hat{\Sigma}_X = 13.36, n_Y = 10, \overline{Y} = 59.90, \hat{\Sigma}_Y = 16.41.$$

Có thể tính ra thống kê T ở đây có giá trị bằng $(69.58-59.90)/6.3415 \approx 1.526$, và số bậc tự do là 10 +12 - 2 = 20. Ta có giá trị P bằng $P = P(|T_{20}| \geq 1.526) \approx 14.3\%$, trong đó T_{20} là ký hiệu biến ngẫu nhiên có phân bố Student T với 20 bậc tự do. Tuy rằng $\overline{X} = 69.58$ chênh lệch với $\overline{Y} = 59.90$ khá nhiều, nhưng mà giá trị P ở đây bằng 14.3% là một con số không đủ nhỏ để có thể loại bỏ giả thuyết H_0. Người nghiên cứu này phải điều tra thêm trước khi có thể kết luận là độ hài lòng về cuộc sống của người trẻ cao hơn người già.

Bài tập 5.14. Giả sử khi khảo sát 30 học sinh nam và 30 học sinh nữ ở một trường học lớn, thấy điểm toán trung bình của 30 học sinh nam là 7.0 với độ lệch chuẩn 1.4, còn của 30 học sinh nữ là 7.4 với độ lệch chuẩn 1.5. Có thể kết luận được rằng học sinh nữ giỏi toán hơn học sinh nam ở trường này được không? (Dùng kiểm định Z).

Bài tập 5.15. Người ta muốn kiểm tra hiệu quả của một chương trình xã hội chăm sóc phụ nữ nhà nghèo đang có thai. Khảo sát trên 50 đứa trẻ sinh ra từ các phụ nữ tham gia chương trình này cho thấy các

đứa trẻ này lúc sinh ra nặng trung bình 3000 gam, với độ lệch chuẩn 410 gam. Để so sánh, người ta khảo sát 50 đứa trẻ sinh ra từ các phụ nữ nhà nghèo không tham gia chương trình, và thấy rằng những đứa trẻ này lúc sinh ra có cân nặng trung bình là 2650 gam với độ lệch chuẩn 425 gam. Kiểm định xem chương trình này có giúp làm trẻ em nhà nghèo đạt cân nặng cao lên khi sinh ra không?

Bài tập 5.16. Một viện dưỡng lão làm thí nghiệm sau: chọn 30 người già ngẫu nhiên trong viện, chia làm 2 nhóm mỗi nhóm 15 người. Cho mỗi người một cái cây cảnh. Yêu cầu những người nhóm đầu tiên hàng ngày chăm sóc cây, còn không yêu cầu những người trong nhóm thứ hai chăm sóc cây. Ghi lại số lần than phiền về sức khỏe của những người trong hai nhóm trong vòng 1 tuần sau khi cho cây. Kết quả là:
Nhóm 1 (được yêu cầu chăm sóc cây): 23, 12, 6, 15, 18, 5, 21, 18, 34, 10, 23, 14, 19, 23, 8.
Nhóm 2 (không yêu cầu chăm sóc cây): 35, 21, 24, 26, 17, 23, 37, 22, 16, 38, 23, 41, 27, 24, 32. Hãy xem việc chăm sóc cây có ảnh hưởng đến số lần than phiền về sức khỏe không. (Dùng kiểm định T).

5.4.4 Kiểm định F so sánh hai độ lệch chuẩn

Nhắc lại rằng, nếu biến ngẫu nhiên X có phân bố normal $\mathcal{N}(\mu, \sigma^2)$, và $\hat{\Sigma}^2$ là hàm phương sai mẫu của X với cỡ thực nghiệm n, thì $n\hat{\Sigma}^2/\sigma^2$ có phân bố χ^2 với $n-1$ bậc tự do. Do đó, có thể dùng các phân bố χ^2 trong việc ước lượng và kiểm định về phương sai và độ lệch chuẩn của X (với giả sử phân bố của X là normal).

Tương tự như vậy, trong trường hợp X và Y là hai biến ngẫu nhiên với phân bố normal, thì để kiểm định so sánh độ lệch chuẩn của X với độ lệch chuẩn của Y, ta có thể dùng các phân bố sau:

Định nghĩa 5.9. *Giả sử χ^2_m và χ^2_n là hai biến ngẫu nhiên độc lập có phân bố χ^2 với m và n bậc tự do tương ứng. Khi đó phân bố xác suất của biến ngẫu nhiên*

$$F_{m,n} = \frac{\chi^2_m/m}{\chi^2_n/n} = \frac{n\chi^2_m}{m\chi^2_n} \tag{5.46}$$

được gọi là **phân bố F** *với m và n bậc tự do.*

Kiểm định dùng phân bố F để so sánh độ lệch chuẩn gọi là **kiểm định F**. Ta giả sử rằng X và Y có phân bố normal với độ lệch chuẩn σ_1 và σ_2 tương ứng. Giả thuyết H_0 là $\sigma_1 = \sigma_2$ Gọi S_1^2 và S_2^2 là các hàm phương sai mẫu hiệu chỉnh của X và Y với cỡ thực nghiệm n_1 và n_2 tương ứng. Nếu $\sigma_1 = \sigma_2$ thì S_1^2/S_2^2 có phân bố F với $n_1 - 1$ và $n_2 - 1$ bậc tự do.

Ghi chú 5.8. Phân bố F được gọi như vậy là theo chữ cái đầu của tên của Ronald Fisher. Kiểm định F chỉ thích hợp khi các phân bố của X và Y là normal hoặc rất gần giống normal (nó không được "robust" lắm, khi phân bố chệch đi khỏi normal).

Ví dụ 5.20. (*Cách đo lường nào chính xác hơn*). Giả sử có hai cách đo hàm lượng chất arsenic trong đất. Mỗi cách được thử 10 lần (cho cùng một chỗ đất), với các kết quả như sau (ppm có nghĩa là parts-per-million, tỷ lệ tính theo phần triệu):

Cách	Trung bình (ppm)	Độ lệch chuẩn mẫu hiệu chỉnh (ppm)
I	7.7	0.8
II	7.9	1.2

Phương pháp nào có độ lệch chuẩn thấp hơn thì được coi là phương pháp đo có độ chính xác cao hơn. Ta muốn kiểm định xem có đủ chứng có để coi rằng phương pháp I chính xác hơn phương pháp II không. Giá trị của thống kê F ở đây là $(0.8)^2/(1.2)^2 = 0.4444$. Các số bậc tự do là $10 - 1 = 9$ và $10 - 1 = 9$. Tra máy tính, ta có $P(F_{9,9} \leq 0.4444) \approx 12\%$, là một con số nhỏ, nhưng chưa đủ nhỏ để loại bỏ giả thuyết H_0 (là hai phương pháp có độ chính xác như nhau), cần thí nghiệm thêm.

Công thức để tính hàm mật độ của các phân bố F như sau. Nó có thể được suy ra từ công thức hàm mật độ cho các phân bố χ^2.

Định lý 5.8. *Phân bố F với m và n bậc tự do có hàm mật độ sau:*

$$f_{m,n}(x) = \begin{cases} 0 & \text{nếu } x \leq 0 \\ c\dfrac{x^{(m/2)-1}}{(mx+n)^{(m+n)/2}} & \text{nếu } x > 0, \end{cases} \tag{5.47}$$

trong đó

$$c = \frac{\Gamma((m+n)/2)m^{m/2}n^{n/2}}{\Gamma(m/2)\Gamma(n/2)}. \tag{5.48}$$

5.5 Kiểm định χ^2

Kiểm định ki bình phương (χ^2 test) là kiểm định thường được dùng để kiểm tra một giả thuyết về tính đúng đắn (goodness-of-fit)

một mô hình xác suất với hữu hạn các sự kiện thành phần $\Omega = \{A_1, A_2, \ldots, A_s\}$. (Khi không gian xác suất là vô hạn, thì người ta chia nó ra theo một phân hoạch hữu hạn để dùng kiểm định này). Giả thuyết H_0 ở đây có thể hiểu là các xác suất $P(A_i)$, $i = 1, \ldots, s$, phải bằng các số p_i nào đó (hoặc thỏa mãn các điều kiện gì đó) cho bởi mô hình. Thay vì kiểm định từng giả thuyết $P(A_i) = p_i$ cho từng sự kiện thành phần (tức là phải làm s kiểm định), ta sẽ làm một kiểm định chung cho toàn bộ mô hình xác suất.

Mẫu thực nghiệm ở đây là một dãy n kết quả, mỗi kết quả có dạng "xảy ra sự kiện A_i". Ta gọi n_i là số lần xảy ra A_i trong mẫu. ($\sum_{i=1}^{s} n_i = n$). Số n_i có thể hiểu là một giá trị thực nghiệm của biến ngẫu nhiên N_i = "số lần xảy ra A_i trong n lần thử nghiệm". Để cho dễ hiểu, ta sẽ phân biệt hai trường hợp: 1) Các xác suất p_i là cố định và được cho trước trong mô hình; 2) Mô hình xác suất phụ thuộc k tham số nào đó (ví dụ như mô hình phân bố Poisson phụ thuộc tham số λ), các tham số đó được ước lượng từ mẫu thực nghiệm , và các xác suất p_i được xác định từ các tham số đó.

5.5.1 Trường hợp mô hình xác suất cố định

Theo định lý Pearson 4.15, khi n đủ lớn , phân phối xác suất của biến ngẫu nhiên

$$\sum_{i=1} \frac{(N_i - P(A_i)n)^2}{P(A_i)n} \tag{5.49}$$

có thể xấp xỉ bằng phân phối χ^2 với $s-1$ bậc tự do, với sai số đủ nhỏ có thể bỏ qua. (Nhắc lại rằng, phân phối χ^2 với $s-1$ bậc tự do là phân phối xác suất của biến ngẫu nhiên $\chi^2_{s-1} = Z_1^2 + \ldots + Z_{s-1}^2$,

tổng bình phương của $s-1$ biến ngẫu nhiên Z_i độc lập có cùng phân bố normal chuẩn tắc $\mathcal{N}(0,1)$). Giá trị thực nghiệm

$$\hat{\chi}^2 = \sum_{i=1}^{s} \frac{(\nu_i - p_i n)^2}{p_i n} \tag{5.50}$$

(khi giả sử rằng $P(A_i) = p_i$ là các số cho trước trong mô hình) có thể coi là một giá trị thực nghiệm của χ^2_{s-1}. Một cách dễ nhớ để viết công thức của thống kê $\hat{\chi}^2$ là:

$$\hat{\chi}^2 = \sum_i \frac{(\text{observed}_i - \text{expected}_i)^2}{\text{expected}_i}, \tag{5.51}$$

trong đó *observed* có nghĩa là các giá trị thực nghiệm, còn *expected* là các giá trị kỳ vọng tương ứng (của số lần xảy ra các sự kiện).

Vì thống kê $\hat{\chi}^2$ là số không âm, và đo độ sai số giữa mô hình phân bố xác suất và phân bố thực nghiệm, nên $\hat{\chi}^2$ càng nhỏ thì chứng tỏ mô hình càng khớp với thực nghiệm. Nhỏ ở đây là nhỏ so với phân phối của χ^2_{s-1}. Bởi vậy, nếu $P(\chi^2_{s-1} \geq \hat{\chi}^2)$ càng cao thì độ tin tưởng của ta vào mô hình (giả thuyết H_0) càng cao. Nếu $P(\chi^2_{s-1} \geq \hat{\chi}^2) > \alpha$, với α là một số cho trước theo qui ước (thông thường người ta lấy $\alpha = 5\%$, nhưng cũng có khi lấy $\alpha = 10\%$ hay 1%) thì giả thuyết H_0 được chấp nhận, còn nếu $P(\chi^2_{s-1} \geq \hat{\chi}^2) < \alpha$ thì người ta chấp nhận đối thuyết H_1, tức là coi rằng mô hình bị sai.

Trong thực tế, để tránh sai số quá cao khi áp dụng định lý Pearson, người ta thường đòi hỏi cỡ n của mẫu phải đủ lớn sao cho $p_i n \geq 10$ với mọi $i = 1, \ldots, s$ (hoặc ít ra là với hầu hết các chỉ số i). Những sự kiện A_i với $p_i n < 10$ là những sự kiện "quá hiếm" để có thể kiểm định xác suất của chúng bằng kiểm định χ^2.

Ví dụ 5.21. Một người chơi tung xúc sắc. Tung một con xúc sắc 120 lần, trong đó có 35 lần hiện lên số 6. Hỏi có sự "thiên vị số 6" (chẳng hạn có sự gian lận, hay quân xúc sắc không cân bằng) ở đây không, hay là số 6 hiện lên nhiều là hoàn toàn do ngẫu nhiên?

Mô hình xác suất ở đây gồm hai sự kiện: $A =$ hiện lên số 6, với xác suất (nếu giả sử không có thiên vị) là $1/6$, và $\overline{A} =$ hiện lên số khác 6, với xác suất $5/6$. Số lần thực nghiệm hiện lên 6 là 35 so với kỳ vọng là $120/6 = 20$, còn số lần hiện lên khác 6 là $120 - 35 = 85$ so với kỳ vọng là 100. Thống kê $\hat{\chi}^2$ ở đây là:

$$\hat{\chi}^2 = \frac{(35-20)^2}{20} + \frac{(85-100)^2}{100} = 13,5.$$

Ta có $P(\chi_1^2 \geq 13,5) < 1\%$. Như vậy giả thuyết H_0 bị loại bỏ, và đối thuyết "số 6 được thiên vị" được chấp thuận.

Tất nhiên, ví dụ trên rất đơn giản, với số bậc tự do là 1, và thay vì làm kiểm định ki bình phương, ta có thể làm kiểm định Z cho xác suất của sự kiện hiện lên số 6, cũng sẽ ra kết quả tương đương. Nhưng nếu thay vì chỉ kiểm định xem số 6 có được thiên vị không, ta muốn kiểm định cùng một lúc tất cả các số của xúc sắc xem có số nào được thiên vị không, thì nói chung sẽ phải dùng đến ki bình phương.

Ví dụ 5.22. Một người tung xúc sắc 120 lần, có 28 lần hiện số 1, 14 lần hiện số 2, 26 lần hiện số 3, 18 lần hiện số 4, 15 lần hiện số 5, 19 lần hiện số 6. Hỏi rằng xúc sắc có "cân bằng" không? Giả thuyết "cân bằng" H_0 ở đây là xác suất hiện lên mỗi số trong mỗi lần tung đều là 1/6. Kỳ vọng số lần hiện ra mỗi số trong 120 lần tung đều là 20. Thống kê $\hat{\chi}^2$ ở đây là: $\hat{\chi}^2 = \frac{(28-20)^2}{20} + \frac{(14-20)^2}{20} + \frac{(26-20)^2}{20} +$

$\frac{(18-20)^2}{20} + \frac{(15-20)^2}{20} + \frac{(19-20)^2}{20} = 8,3.$

Tra bảng phân phối xác suất của χ^2_5, ta có $P(\chi^2_5 \geq 8,3) \approx 14\%$. Con số 14% đủ lớn để chấp nhận giả thuyết H_0.

5.5.2 Trường hợp mô hình xác suất được ước lượng theo tham số

Nhắc lại rằng, khi các xác suất $p_i = P(A_i)$, $i = 1, \ldots, s$ là được cho trước trong mô hình và ta cần kiểm định chúng, thì số bậc tự do của phân bố χ^2 tương ứng là $s - 1$. Lý do là vì ta có một ràng buộc tuyến tính giữa s biến ngẫu nhiên $\frac{N_i - p_i n}{\sqrt{p_i(1-p_i)n}}$, cụ thể là:

$$\sum_{i=1}^{s} \sqrt{p_i(1-p_i)} \frac{N_i - p_i n}{\sqrt{p_i(1-p_i)n}} = \sum_i \frac{N_i - p_i n}{\sqrt{n}} = 0. \qquad (5.52)$$

Giới hạn của phân bố xác suất đồng thời của bộ s biến ngẫu nhiên này là một phân bố normal s chiều nhưng có rank bằng $s-1$ vì có một ràng buộc tuyến tính, nên nó có thể nhận được từ phân bố normal chuẩn tắc $s-1$ chiều qua một phép biến đổi tuyến tính, và bởi vậy ta chỉ có $s-1$ bậc tự do.

Khi các xác suất $p_i = P(A_i)$, $i = 1, \ldots, s$ *không được cho trước* trong mô hình, mà phụ thuộc vào k tham số $\theta_1, \ldots, \theta_k$ nào đó của mô hình phân bố xác suất, và k tham số này được ước lượng từ mẫu thực nghiệm, thì thay vì 1 ràng buộc tuyến tính, ta có $k+1$ ràng buộc tuyến tính giữa các biến ngẫu nhiên $\frac{N_i - p_i n}{\sqrt{p_i(1-p_i)n}}$. Bởi vậy, trong trường hợp này, phân phối xác suất của $\sum_{i=1} \frac{(N_i - p_i n)^2}{p_i n}$ (với $k+1$

điều kiện ràng buộc đó) không tiến tới phân phối xác suất của χ^2_{s-1} (với $s-1$ bậc tự do) nữa, mà tiến tới phân phối xác suất của χ^2_{s-k-1} (với $s-k-1$ bậc tự do). Khẳng định này có thể được chứng minh tương tự như định lý Pearson 4.15. Bởi vậy, trường hợp khi mà mô hình xác suất có k tham số được ước lượng, được kiểm định hoàn toàn tương tự như trường hợp không có tham số, nhưng ở bước cuối cùng phải dùng phân phối xác suất χ^2 với $s-k-1$ bậc tự do thay vì $s-1$ bậc tự do: Nếu $P(\chi^2_{s-k-1} \geq \hat{\chi}^2) > \alpha$ thì giả thuyết H_0 được chấp nhận, còn nếu $P(\chi^2_{s-k-1} \geq \hat{\chi}^2) < \alpha$ thì chấp nhận đối thuyết H_1.

Ví dụ 5.23. Chúng ta sẽ kiểm định giả thuyết "số vụ án mạng xảy ra ở London hàng ngày tuân theo phân bố Poisson", dựa theo số liệu thống kê trong ví dụ 5.4. Ta có bảng thống kê sau:

i	0	1	2	3	4
n_i	713	299	66	16	1

trong đó n_i là số ngày xảy ra i án mạng trong vòng 3 năm, từ 04/2004 đến 03/2007. Tổng số ngày ở đây là $713 + 299 + 66 + 16 + 1 = 1095$ ngày.

Trước hết ta ước lượng tham số λ của phân bố Poisson trong giả thuyết. Nếu X là biến ngẫu nhiên tuân theo phân phối Poisson với tham số λ, thì $\lambda = \mathbb{E}(X)$. Bởi vậy ta ước lượng λ bằng kỳ vọng của mẫu thực nghiệm:

$$\lambda = \frac{\sum_{i=0}^{4} in_i}{\sum_{i=0}^{i} n_i} = \frac{1}{1095}(0\times713+1\times299+2\times66+3\times16+4\times1) \approx 0,4411.$$

Gọi p_{0i} là xác suất của sự kiện "trong ngày có i vụ giết người" theo

mô hình phân phối Poisson với tham số λ. Khi đó

$$p_{0i} = (0,4411)^i e^{-0,4411}/i!, \quad i = 0, 1, 2, \ldots$$

Việc ước lượng λ (và qua đó p_{0i}) như trên tạo thêm một ràng buộc tuyến tính sau đây cho các biến ngẫu nhiên $\dfrac{N_i - p_{0i}n}{\sqrt{p_{0i}(1-p_{0i})n}}$, ngoài ràng buộc cho bởi phương trình (5.52), cụ thể là:

$$\sum_i i \frac{N_i - p_{0i}n}{\sqrt{n}} = 0.$$

Nhân $n = 1095$ với p_{0i}, ta được bảng sau, với các giá trị kỳ vọng np_{0i} về số ngày có i vụ giết người trong vòng 3 năm: ta được bảng sau:

i	0	1	2	≥ 3
np_{0i}	704,44	310,73	68,53	11,28

Ở bảng trên, ta gộp các số các số $np_{0i}, i \geq 3$, lại với nhau, để được một giá trị lớn hơn 10 (các số bắt đầu từ np_{04} trở đi quá nhỏ: $np_{04} \approx 1,1$, $np_{05} \approx 0,1$, $np_{06} < 0,01$, ...). Tức là ta sẽ kiểm định mô hình phân bố Poisson đơn giản hóa, với chỉ có 4 sự kiện thành phần, ứng với số vụ giết người trong ngày như sau: $0, 1, 2, \geq 3$.

Giá trị của thống kê $\hat{\chi}^2$ ở đây là:

$$\hat{\chi}^2 \approx \frac{(713-704,44)^2}{704,44} + \frac{(299-310,73)^2}{310,73} + \frac{(66-68,53)^2}{68,53} + \frac{(17-11,28)^2}{11,28} \approx 3,54.$$

Vì trong mô hình phân bố xác suất có $k = 1$ tham số được tính bằng ước lượng (tham số λ), nên phân bố xác suất χ^2 cần dùng ở đây

có số bậc tự do bằng $4 - 1 - 1 = 2$. Ta có

$$P(\chi_2^2 \geq \hat{\chi}^2) \approx P(\chi_2^2 \geq 3,54) \approx 17\%,$$

là con số khá lớn (lớn hơn 10%). Bởi vậy giả thuyết H_0 (rằng số vụ án mạng hàng ngày tuân theo phân bố Poisson) được chấp nhận.

5.5.3 Kiểm định χ^2 cho sự độc lập

Khi ta muốn kiểm tra xem hai sự kiện hay hai biến ngẫu nhiên nào đó có độc lập với nhau không, ta cũng có thể dùng χ^2. Chẳng hạn, giả sử ta có biến ngẫu nhiên X nhận m giá trị $x_1, \ldots, x_m$, và biến ngẫu nhiên Y nhận n giá trị $y_1, \ldots, y_n$. Giả thuyết H_0: X độc lập với Y có nghĩa là $P(i,j) := P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$ với mọi i, j. Mô hình gian xác suất ở đây có mn phần tử $(X = x_i, Y = y_j)$. Mô hình xác suất ở đây có $m + n - 2$ tham số, có nghĩa là nếu ta ước lượng được $m + n - 2$ giá trị $P(X = x_1), \ldots, P(X = x_{m-1}), P(Y = y_1), \ldots, P(Y = y_{n-1})$, thì ta biết được toàn bộ phân bố xác suất của không gian xác suất (nếu chấp nhận giả thuyết $P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$ với mọi i, j). Bởi vậy, số bậc tự do của phân bố χ^2 cần dùng trong kiểm định giả thuyết H_0 ở đây là: $mn - (m + n - 2) - 1 = (m - 1)(n - 1)$.

Ví dụ 5.24. Người ta muốn kiểm định xem độ tuổi của người có ảnh hưởng đến khuynh hướng chính trị không. Để đơn giản, trong ví dụ này ta chia các khuynh hướng chính trị ra làm 3 khuynh hướng: phái tả, phái hữu, và trung lập. Và ta cũng chia các độ tuổi ra làm 3: dưới 30 tuổi, từ 30 đến 50, và trên 50 tuổi. Bảng sau là một bảng thống

kê thăm dò khuynh hướng của 500 người được chọn một cách ngẫu nhiên:

Tuổi Khuynh hướng	Phái tả	Phái hữu	Trung lập	Tổng
Dưới 30	45	35	38	118
30 đến 50	62	60	95	217
Trên 50	48	49	68	165
Tổng	155	144	201	500

Giả sử độ tuổi và khuynh hướng chính trị độc lập với nhau. Khi đó, dựa vào các số tổng trong bảng trên, ước lượng kỳ vọng của số người dưới 30 tuổi theo phái tả trong số 500 người sẽ là $155 \times 118/500 = 36,58$, và tương tự như vậy cho các ô khác. Thống kê $\hat{\chi}^2$ ở đây là:

$$\begin{aligned}\hat{\chi}^2 = & \frac{(45-36,58)^2}{36,58} + \frac{(62-67,27)^2}{67,27} + \frac{(48-51,15)^2}{51,15} \\ & + \frac{(35-33,984)^2}{33,984} + \frac{(60-62,496)^2}{62,496} + \frac{(49-47,52)^2}{47,52} \\ & + \frac{(38-47,436)^2}{47,436} + \frac{(95-87,23)^2}{87,23} + \frac{(68-66,33)^2}{66,33} \approx 5,329.\end{aligned}$$

Số bậc tự do ở đây là $(m-1)(n-1) = (3-1)(3-1) = 4$. Ta có $P(\chi_4^2 \geq 5,329) > 25\%$. Như vậy ta chấp nhận giả thiết H_0: độ tuổi không ảnh hưởng (đáng kể) tới khuynh hướng chính trị.

Bài tập 5.17. (*Sinh viên nữ có bị kỳ thị?*). Một điều tra năm 1975 ở một trường đại học hàng đầu trên thế giới về số sinh viên nam và nữ xin học và được nhận vào học các chương trình sau đại học ở 3 khoa lớn nhất trường cho kết quả thống kê sau:

	Được nhận	Bị từ chối
Nam	526	550
Nữ	313	698

Hãy kiểm định xem có đủ cơ sở thống kê để nói rằng sinh viên nữ khó được nhận vào học sau đại học hơn so với sinh viên nam không?

5.6 Phân tích hồi qui

Hồi qui (regression) là phương pháp thống kê toán học để ước lượng và kiểm định các quan hệ giữa các biến ngẫu nhiên, và có thể từ đó đưa ra các dự báo. Các quan hệ ở đây được viết dưới dạng các hàm số hay phương trình.

Ý tưởng chung như sau: giả sử ta có một biến ngẫu nhiên Y, mà ta muốn ước lượng xấp xỉ dưới dạng một hàm số $F(X_1, \ldots, X_s)$ của các biến ngẫu nhiên $X_1, \ldots, X_s$ khác (gọi là các **biến điều khiển** (control variables), hay còn gọi là **biến tự do** (tiếng Anh gọi là independent variables, nhưng không có nghĩa đây là một bộ biến ngẫu nhiên độc lập), trong khi Y được gọi là **biến phụ thuộc** (dependent variable)), tức là khi ta có các giá trị của $X_1, \ldots, X_s$, thì ta muốn từ đó ước lượng được giá trị của Y. Hàm số F này có thể phụ thuộc vào một số tham số $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ nào đó. Ta có thể viết Y như sau:

$$Y = F_{\boldsymbol{\theta}}(X_1, \ldots, X_s) + \epsilon, \tag{5.53}$$

trong đó ϵ là phần sai số (cũng là một biến ngẫu nhiên). Ta muốn chọn hàm F một cách thích hợp nhất có thể (phụ thuộc vào từng lớp

bài toán cụ thể), và các tham số $\boldsymbol{\theta}$, sao cho sai số ϵ là nhỏ nhất có thể. Thông thường, người ta đo độ to nhỏ của sai số bằng chuẩn L_2 (sai số trung bình bình phương). Có nghĩa là, ta muốn chọn $\boldsymbol{\theta}$ sao cho $\mathbb{E}(|\epsilon|^2)$ là nhỏ nhất có thể. Đại lượng

$$\sqrt{\mathbb{E}(|\epsilon|^2)} \tag{5.54}$$

được gọi là **sai số chuẩn** (standard error) của mô hình hồi qui. Mô hình nào mà có sai số chuẩn càng thấp thì được coi là càng chính xác.

Mô hình đơn giản nhất là mô hình tuyến tính với một biến điều khiển: $F(X) = aX + b$, với a và b là hằng số. Việc tìm a, b rồi ước lượng Y bởi hàm tuyến tính $aX + b$ được gọi là **hồi qui tuyến tính đơn**, mà ta đã gặp trong Chương 3, Mục 3.4.3. Hồi qui tuyến tính thích hợp trong một số trường hợp, khi các biến ngẫu nhiên phải có quan hệ tuyến tính nào đó với nhau về mặt lý thuyết. Chẳng hạn, sự phụ thuộc của giá nhà vào diện tích nhà (không kể đến các yếu tố khác) có thể coi là tuyến tính, vì ta có thể hình dung là 1 cái nhà to có thể chia làm hai cái nhà nhỏ bằng một nửa. Thế nhưng trọng lượng của quả táo không phụ thuộc tuyến tính vào đường kính của nó, mà phụ thuộc tuyến tính vào lập phương của đường kính của nó thì hợp lý hơn. Hay dân số của Việt Nam thay đổi hàng năm cũng không theo kiểu tuyến tính. Bởi vậy, việc chọn lựa hàm F sao cho thích hợp (dựa trên các lý thuyết nào đó) là quan trọng khi áp dụng phương pháp hồi qui. Một khi đã cố định một lớp hàm $F_{\boldsymbol{\theta}}$ hợp lý, giá trị của $\boldsymbol{\theta}$ hợp lý nhất sẽ là giá trị sao cho $\mathbb{E}((Y - F_{\boldsymbol{\theta}}(X_1, \ldots, X_s))^2)$ là nhỏ nhất. Như vậy, trong nhiều trường hợp, bài toán hồi qui được đưa về vấn đề tìm cực trị: tìm $\boldsymbol{\theta}$ sao cho $\mathbb{E}((Y - F_{\boldsymbol{\theta}}(X_1, \ldots, X_s))^2)$ nhỏ nhất.

Nhắc lại rằng, trong thực tế, vì ta không biết chính xác phân bố

xác suất đồng thời của các biến X_i và Y, mà chỉ biết một phân bố thực nghiệm nào đó thông qua một số số liệu kết quả thực nghiệm, nên ta sẽ thay (ước lượng) các không gian xác suất bởi các không gian xác suất thực nghiệm.

5.6.1 Hồi qui tuyến tính đơn

Hồi qui tuyến tính đơn đã được nhắc tới trong Mục 3.4.3. Giả sử hai biến ngẫu nhiên X, Y hợp thành vector ngẫu nhiên 2 chiều (X, Y), với các giá trị thực nghiệm $(x_1, y_1), \ldots, (x_n, y_n)$. Ta muốn viết Y dưới dạng hàm tuyến tính của X,

$$Y = aX + b + \epsilon \tag{5.55}$$

với sai số bình phương $\mathbb{E}(|\epsilon|^2)$ nhỏ nhất (a, b là hằng số còn sai số ϵ là biến ngẫu nhiên). Ta sẽ tìm a, b sao cho sai số thực nghiệm bìng phương (sai số trung bình bình phương) $\sum_{i=1}^{n} |\epsilon_i|^2/n$ là nhỏ nhất, trong đó $\epsilon_i = y_i - ax_i - b$ là các sai số thực nghiệm. Gọi vector với phân bố thực nghiệm cho bởi các cặp số $(x_1, y_1), \ldots, (x_n, y_n)$ này là $(\tilde{X}, \tilde{Y})$. Khi đó công thức sẽ là (xem Mục 3.4.3):

$$a = \frac{cov(\tilde{X}, \tilde{Y})}{var(\tilde{X})} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2} \tag{5.56}$$

và

$$b = \overline{y} - a\overline{x} = \mathbb{E}(\tilde{Y}) - a\mathbb{E}(\tilde{X}), \tag{5.57}$$

trong đó $\overline{x} = (\sum x_i)/n$ và $\overline{y} = (\sum y_i)/n$ là các giá trị kỳ vọng thực nghiệm. Bình phương của hệ số tương quan thực nghiệm,

$$R^2 = \frac{cov(\tilde{X}, \tilde{Y})^2}{var(\tilde{X}) var(\tilde{Y})} \tag{5.58}$$

là số đo độ chính xác của hồi qui tuyến tính trên mẫu thực nghiệm: nếu $R^2 = 1$ thì $\sum_{i=1}^n |\epsilon_i|^2 = 0$, tức là không có sai số. Trong trường hợp tổng quát, ta có

$$R^2 = \frac{var(a\tilde{X} + b)}{var(\tilde{Y})} = \frac{var(\tilde{Y}) - (1/n)\sum_i \epsilon_i^2}{var(\tilde{Y})}, \tag{5.59}$$

và tức là R^2 càng gần 1, thì tổng sai số bình phương $\sum_i \epsilon_i^2$ càng nhỏ. Ví dụ, khi $R^2 = 0.9$, thì độ sai số chuẩn (căn bậc hai của sai số trung bình bình phương) bằng $\sqrt{1 - R^2} \approx 0.32$ lần độ lệch chuẩn (thực nghiệm) của Y. Nếu giả sử độ lệch chuẩn của Y bằng $1/4$ giá trị trung bình của Y, thì tức là hồi qui tuyến tính trong trường hợp này sẽ có sai số ϵ vào quãng $32\%/4 = 8\%$ giá trị của Y.

Khi đã có phương trình hồi qui tuyến tính đơn $Y = aX + b + \epsilon$, thì với mỗi giá trị của X ta có một ước lượng $\hat{Y}$ cho giá trị tương ứng của Y theo công thức $\hat{Y} = aX + b$. Giống như các bài toán ước lượng được bàn ở phía trước, có thể tính khoảng tin cậy và độ tin cậy của ước lượng này. Khi R^2 gần 1, và X nằm trong đoạn thẳng $[\min x_i, \max x_i]$ thì ước lượng này có độ chính xác cao (khoảng tin cậy hẹp) còn ngược lại thì độ chính xác thấp. Chúng ta sẽ không đi vào chi tiết ở đây.

5.6.2 Hồi qui tuyến tính bội

Trong **hồi qui tuyến tính bội**, ta muốn tìm tham số $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_i)$, sao cho khi đặt

$$Y = \theta_0 + \sum_{i=1}^{s} \theta_i X_i + \epsilon, \tag{5.60}$$

trong đó X_i và Y là các biến ngẫu nhiên cho trước, thì sai số trung bình bình phương $\mathbb{E}(|\epsilon|^2)$ là nhỏ nhất.

Để cho tiện, ta sẽ đặt $X_0 = 1$ và coi đó như là một biến ngẫu nhiên (có giá trị luôn bằng 1), và viết

$$Y = \sum_{i=0}^{s} \theta_i X_i + \epsilon. \tag{5.61}$$

Ta sẽ giả sử rằng các biến ngẫu nhiên $X_i, i = 0, \ldots, s$ là độc lập tuyến tính với nhau (không có biến nào có thể viết được dưới dạng một tổ hợp tuyến tính của các biến khác), vì nếu chúng phụ thuộc tuyến tính, thì ta có thể loại bớt một số biến đi.

Không gian các biến ngẫu nhiên (trên cùng một không gian xác suất ban đầu) với tích vô hướng $\langle X, Y\rangle := \mathbb{E}(XY)$ là một không gian (tiền) Hilbert. Bởi vậy biến ngẫu nhiên $\tilde{Y} = \sum_{i=0}^{s} \theta_i X_i$ nằm trên không gian con $s+1$ chiều $V = \mathbb{R}\langle X_0, \ldots, X_s\rangle$ sinh bởi $X_0, \ldots X_s$, sao cho chuẩn bình phương $\|Y - \tilde{Y}\|^2 := \mathbb{E}(|Y - \tilde{Y}|^2)$ nhỏ nhất, chính là ảnh của phép chiếu vuông góc từ Y lên trên không gian con V này. Nói cách khác, các tham số θ_i cần thỏa mãn hệ phương trình tuyến tính sau:

$$\langle Y - \sum_{i=0}^{s} \theta_i X_i, X_j\rangle = 0 \ \ \forall\, j = 0, 1, \ldots, s, \tag{5.62}$$

hay có thể viết là

$$\sum_{i=0}^{s} \theta_i \langle X_i, X_j\rangle = \langle Y, X_j\rangle \ \ \forall\, j = 0, 1, \ldots, s. \tag{5.63}$$

Nghiệm duy nhất của hệ phương trình này là:

$$(\theta_i)_{i=0,\ldots,s} = \left((\langle X_i, X_j\rangle)_{i=0,\ldots,s}^{j=0,\ldots,s} \right)^{-1} .(\langle Y, X_j\rangle)_{j=0,\ldots,s}. \tag{5.64}$$

5.6.3 Hồi qui phi tuyến

Hồi qui phi tuyến là khi hàm hồi qui F không phải là hàm tuyến tính của các biến X_i. Tuy nhiên, trong nhiều trường hợp, bằng cách đổi biến, ta có thể đưa bài toán hồi qui phi tuyến về bài toán hồi qui tuyến tính bội. Ví dụ, giả sử hàm F là hàm đa thức bậc 3 một biến: $F(X) = aX^3 + bX^2 + cX + d$. Khi đó, đặt $X_1 = X, X_2 = X^2, X_3 = X^3$, ta đưa bài toán này về trường hợp hồi qui tuyến tính với ba biến điều khiển X_1, X_2, X_3. Các biến điều khiển này tất nhiên là phụ thuộc vào nhau, nhưng chúng độc lập tuyến tính với nhau, bởi vậy có thể dùng nguyên tắc giải bài toán hồi qui tuyến tính bội như trong mục phía trên. Trong trường hợp chung, khi mà không đưa được về mô hình tuyến tính, việc tính toán có thể phức tạp hơn, nhưng các chương trình máy tính sẽ giúp chúng ta tìm được các tham số tốt nhất, và kiểm tra mức độ sai số của mô hình.

Ví dụ 5.25. Chúng ta sẽ thử áp dụng một số mô hình hồi qui vào việc lước lượng giá của các xe ô tô BMW 320 cũ bán ở Pháp vào 11/2009. Gía của xe phụ thuộc vào nhiều yếu tố: tuổi của xe, số km đã chạy, kiểu dáng xe, tiện nghi trong xe, sự bảo hành, các phụ tùng đã thay thế, v.v. Ở đây, để đơn giản, ta sẽ chỉ ước lượng giá xe theo hai biến: tuổi của xe và số km đã chạy. Tất nhiên ước lượng như vậy sẽ có sai số cao, và muốn ước lượng chính xác hơn phải thêm các biến khác. Bảng sau đây là giá bán (tính theo nghìn euro) của 60 chiếc BMW cũ tại thời điểm 08/11/2009, cùng với tuổi của xe (tính theo số năm) và quãng đường đã chạy (tính theo nghìn km):

```
Obs        price        age      distance
```

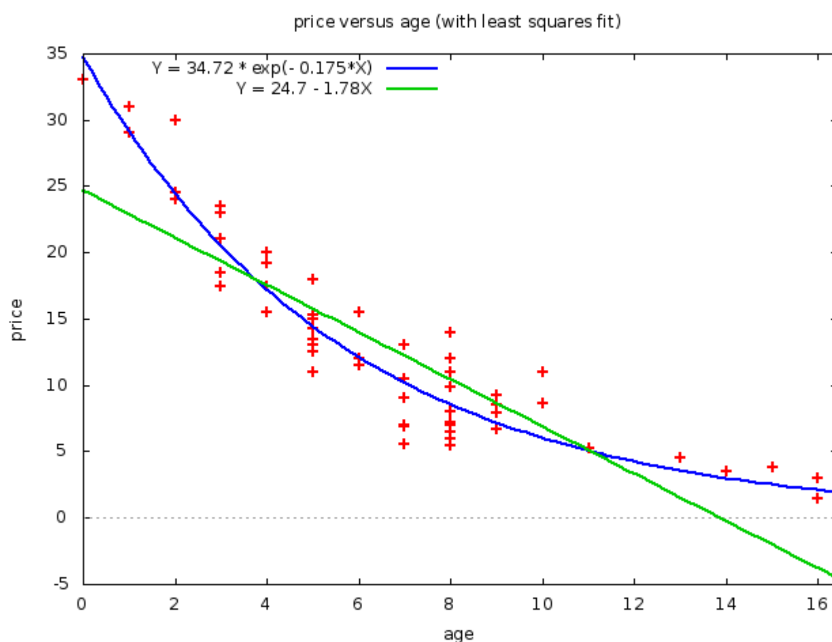
1	31.0	1	24
2	12.5	5	115
3	15.5	6	80
4	6.7	9	195
5	30.0	2	53
6	21.0	3	52
7	18.5	3	75
8	8.6	10	126
9	9.0	7	138
10	18.0	5	70
11	11.0	5	150
12	13.0	5	156
13	11.0	8	124
14	9.0	7	180
15	8.0	8	143
16	12.0	8	97
17	17.5	4	100
18	7.0	8	200
19	20.0	4	80
20	6.0	8	230
21	15.3	5	109
22	23.0	3	37
23	4.5	13	130
24	7.0	8	180
25	24.5	2	25
26	12.5	5	142
27	15.0	5	70
28	7.0	7	166
29	24.0	2	45
30	11.5	6	146
31	23.5	3	55
32	7.2	8	245
33	29.0	1	13
34	9.9	8	188
35	33.0	0	10
36	14.3	5	90
37	17.5	3	101

38	12.0	6	116
39	6.5	8	182
40	5.6	7	223
41	11.0	5	124
42	17.5	3	101
43	13.5	5	73
44	19.2	4	61
45	6.9	7	216
46	1.5	16	246
47	13.0	7	135
48	11.0	10	105
49	9.3	9	145
50	3.8	15	78
51	13.0	5	119
52	14.0	8	86
53	15.5	4	73
54	10.5	7	130
55	8.5	9	161
56	3.5	14	175
57	3.0	16	165
58	7.9	9	126
59	5.5	8	258
60	5.3	11	273

Mô hình thứ nhất là mô hình hồi qui tuyến tính đơn, của giá theo tuổi: price $= a + b.$age. Máy tính cho kết quả sau:

$$\text{estimated_price} \approx 24.69 - 1.78 \times \text{age}, \tag{5.65}$$

với sai số chuẩn bằng 3.78 (so với giá trung bình của xe là 13.03). Sai số chuẩn như vậy là rất cao so với giá trung bình ($3.78/13.03 \approx 29\%$). Rõ ràng mô hình này không được tốt, vì chẳng hạn nó cho ước lượng giá âm cho những xe trên 14 tuổi, trong khi những xe đó vẫn có giá dương mấy nghìn euro.



Hình 5.6: Mô hình hồi qui tuyến tính đơn và phi tuyến đơn cho giá xe BMW

Mô hình thứ hai là mô hình phi tuyến đơn: price $= a.\exp(b.$age$)$. Theo mô hình này, giá của xe giảm theo tuổi, không theo cấp số cộng mà theo cấp số nhân. Máy tính cho kết quả sau:

$$\text{estimated_price} \approx 34.72 \times \exp(-0.175 \times \text{age}), \tag{5.66}$$

với sai số chuẩn là 2.36. Sai số chuẩn này đã giảm đáng kể so với mô hình tuyến tính (từ 3.78 xuống còn 2.36), và hơn nữa mô hình này hợp lý hơn về mặt logic, vì giá của xe được đem bán luôn là số dương

(xe nào hết giá trị, thì người ta vứt vào bãi thải xe, không còn đem bán nữa).

Mô hình thứ ba là tuyến tính đơn theo quãng đường đã chạy: price $= a + b.$distance. Máy tính cho kết quả:

$$\text{estimated_price} \approx 25.05 - 0.096 \times \text{distance}, \tag{5.67}$$

với sai số chuẩn là 4.08. Mô hình này còn tồi hơn là mô hình hồi qui tuyến tính đơn theo biến tuổi của xe.

Mô hình thứ tư là tuyến tính bội, theo tuổi của xe và quãng đường đã chay. Máy tính cho kết quả:

$$\text{estimated_price} \approx 27.50 - 0.0557 \times \text{distance} - 1.146 \times \text{age}, \tag{5.68}$$

với sai số chuẩn là 2.60. Mô hình này tất nhiên tốt hơn cả hai mô hình hồi qui tuyến tính đơn phía trên, nhưng sai số của nó vẫn cao hơn là mô hình phi tuyến đơn. Lý do khá hiển nhiên: sự phụ thuộc của giá xe vào tuổi là phi tuyến.

Mô hình thứ năm là kết hợp của mô hình thứ hai và thứ ba: phi tuyến theo tuổi cộng thêm một phần tuyến tính theo quãng đường đã chạy. Máy tính cho kết quả sau:

$$\text{estimated_price} \approx 31.58 \times \exp(-0.1075 \times \text{age}) - 0.0297 \times \text{distance}, \tag{5.69}$$

với sai số chuẩn là 2.07. Mô hình này chính xác hơn cả 4 mô hình phía trước.

Mô hình thứ sáu là điều chỉnh của mô hình thứ năm. Ta sẽ thay biến quãng đường đã chạy bằng một biến mới, gọi là attrition (hao

mòn):

$$\text{attrition} = \frac{\text{distance}}{(\text{age} + 0.5)} - 10. \tag{5.70}$$

Ý tưởng là, các xe nói chung chạy ít ra 10 nghìn km một năm. Mức 10 nghìn km một nãm được coi là mức với độ hao mòn thấp, và với độ hao mòn đó thì giá xe giảm theo cấp số nhân. Nếu chạy trên 10 nghìn km một năm, thì độ hao mòn cao hơn mức thấp, và giá của xe giảm thêm đi. Máy tính cho kết quả:

$$\text{estimated_price} \approx 35.83 \times \exp(-0.1468 \times \text{age}) - 0.2815 \times \text{attrition}, \tag{5.71}$$

với sai số chuẩn là 1.70, tốt hơn nhiều so với các mô hình trước. Có thể xây dựng thêm những mô hình khác hợp lý và chính xác hơn nữa, nhưng chúng ta sẽ tạm dừng ở đây.

Bài tập 5.18. Hãy lấy những bảng số liệu thống kê có thực bất kỳ nào đó (chẳng hạn như những bảng số liệu thống kê đi kèm theo chương trình gretl, hoặc là những bảng số liệu thống kê từ rất nhiều nguồn khác nhau trên internet), rồi thử làm phân tích hồi quy tuyến tính đơn, tuyến tính bội, và phi tuyến với chúng.

Phụ lục A

Lời giải cho một số bài tập

1.1 Lời giải bài tập Chương 1

Bài tập 1.1. $3' \Rightarrow 3$: Hiển nhiên.

$3 \Rightarrow 3'$: Đặt $B' = (A \cup B) \setminus A$. Khi đó $A \cap B' = \varnothing$ và $A \cup B = A \cup B'$. Suy ra $P(A \cup B) = P(A \cup B') = P(A) + P(B')$. Kết hợp với $P(B) = P(B') + P(A \cap B)$ ta được $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Bài tập 1.2. Áp dụng tiên đề 3' ta có $P(A \cup B \cup C) = P(A \cup B) + P(C) - P((A \cup B) \cap C)$
$= P(A) + P(B) - P(A \cap B) + P(C) - P((A \cap C) \cup (B \cap C))$
$= P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P((A \cap C) \cap (B \cap C))$
$= P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$.

Bằng quy nạp ta có thể chứng minh được rằng

$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i\neq j} P(A_i \cap A_j) + \sum_{i\neq j\neq k} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1} P(\bigcap_{i=1}^{n} A_i)$.

Bài tập 1.3. Có $n!$ cách xếp n bạn thành một hàng dọc, trong đó có $(n-1)!$ cách xếp để Vôva ở ngay sau Lily. Như vậy xác suất để Vôva ở ngay sau Lily trong hàng là $\frac{(n-1)!}{n!} = \frac{1}{n}$.

Có thể giải cách khác như sau. Xác suất để Vôva không đứng đầu hàng là $(n-1)/n$. Khi Vôva đứng đầu hàng thì không thể đứng sang Lily, còn khi Vôva không đứng đầu hàng, thì xác suất để Lily đứng ngay trước Vôva là $1/(n-1)$ (vì trong $n-1$ vị trí còn lại thì có 1 vị trí là ngay trước Vôva). Bởi vậy xác suất để Vôva đứng ngay sau Lily là: $(n-1)/n \times 1/(n-1) = 1/n$.

Bài tập 1.4. Gọi Ω là không gian mẫu, A là biến cố có hai người trong nhóm viết tên của nhau. Ta có $|\Omega| = 4^5 = 1024$, $|A| = C_5^2.4^3 - C_5^4.C_4^2.4 = 520$. (Đầu tiên chọn ra hai người trong nhóm viết tên của nhau, 3 người còn lại viết tên một người bất kỳ trong nhóm, như vậy những cách viết tên mà có hai cặp trong nhóm viết tên của nhau đã được tính hai lần). Xác suất để có hai người trong nhóm viết tên của nhau là $P(A) = \frac{|A|}{|\Omega|} = \frac{65}{128}$.

Bài tập 1.5. Gọi Ω là không gian mẫu, M là biến cố đội A gặp đội B trong giải, M_1, M_2, M_3 lần lượt là các biến cố đội A gặp đội B ở vòng 1, vòng 2, vòng 3. Ta có $P(M) = P(M_1) + P(M_2) + P(M_3)$, với $P(M_1) = \frac{1}{7}$. $|\Omega| = C_8^2.6! = 28.6!$. (Có C_8^2 cách chọn hai đội ở vòng chung kết, 6.5 cách chọn hai đội bị thua ở vòng 2 và $4.3.2.1$ cách

chọn 4 đội bị thua ở vòng 1). $|M_3| = 6!$ (Có 6.5 cách chọn hai đội bị thua A và B ở vòng 2 và $4.3.2.1$ cách chọn 4 đội bị thua ở vòng 1). $|M_2| = 2.6!$ (Có 2 cách chọn đội thắng trong trận $A - B$, 6 cách chọn đội gặp A hoặc B trong trận chung kết, 5 cách chọn đội thứ hai bị thua ở vòng 2 và $4.3.2.1$ cách chọn 4 đội bị thua ở vòng 1). Như vậy $P(M_2) = \dfrac{|M_2|}{|\Omega|} = \dfrac{1}{14}$, $P(M_3) = \dfrac{|M_3|}{|\Omega|} = \dfrac{1}{28}$, $P(M) = \dfrac{1}{7} + \dfrac{1}{14} + \dfrac{1}{28} = \dfrac{1}{4}$.

Có thể giải cách khác như sau. Tổng cộng có $C_8^2 = 28$ cặp đội, và có 7 trận đấu. Vì các cặp là "bình đẳng", nên trung bình mỗi cặp có $7/28 = 1/4$ trận đấu (giữa hai đội của cặp đó). Có nghĩa là cặp đội $A - B$ cũng có trung bình là 1/4 trận đấu, hay nói cách khác, xác suất để xảy ra trận đấu giữa A và B là 1/4.

Bài tập 1.6. Tính phản xạ và tính đối xứng là hiển nhiên. Ta chứng minh tính chất bắc cầu. Giả sử $\phi : (\Omega_1, P_1) \longrightarrow (\Omega_2, P_2)$ và $\psi : (\Omega_2, P_2) \longrightarrow (\Omega_3, P_3)$ là các đẳng cấu xác suất với $\phi : \Omega_1 \backslash A_1 \longrightarrow \Omega_2 \backslash A_2$ và $\psi : \Omega_2 \backslash B_2 \longrightarrow \Omega_3 \backslash B_3$ là các song ánh bảo toàn xác suất, $P_1(A_1) = P_2(A_2) = P_2(B_2) = P_3(B_3) = 0$. Đặt $A_3 = \psi(A_2), B_1 = \phi^{-1}(B_2)$. Khi đó $P_1(B_1) = P_3(A_3) = 0$. Ánh xạ $\psi \circ \phi : \Omega_1 \backslash (A_1 \cup B_1) \longrightarrow \Omega_3 \backslash (A_3 \cup B_3)$ là một song ánh bảo toàn xác suất với $P(A_1 \cup B_1) = P(A_3 \cup B_3) = 0$. Vậy (Ω_1, P_1) và (Ω_3, P_3) đẳng cấu xác suất.

Bài tập 1.7. Xét ánh xạ chiếu $\phi_1 : (\Omega_1, P_1) \times (\Omega_2, P_2) \longrightarrow (\Omega_1, P_1)$ và A là một tập P_1 - đo được. Khi đó $\phi_1^{-1}(A) = A \times \Omega_2$ là P - đo được và $P(\phi_1^{-1}(A)) = P(A \times \Omega_2) = P_1(A).P_2(\Omega_2) = P_1(A)$.

Bài tập 1.8. Để kết thúc trận đấu, Nam và Tiến phải chơi ít nhất là 3 sét và nhiều nhất là 5 sét. Để Nam là người thắng trận thì Nam

phải là người thắng set cuối cùng. Xác suất để Nam thắng trận là $C_2^2.(\frac{2}{5})^3 + C_3^2.(\frac{2}{5})^3.\frac{3}{5} + C_4^2.(\frac{2}{5})^3.(\frac{3}{5})^2 = 0.31744$.

Bài tập 1.9. $P(A|B) = P(B|A) \Leftrightarrow \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(A)} \Leftrightarrow P(A) = P(B)$.

Bài tập 1.10. Gọi A là biến cố trong 3 con mèo có ít nhất một con là mèo cái, B là biến cố cả 3 con mèo đều là mèo cái. Ta cần tính $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B)}{P(A)} = \frac{(1/2)^3}{1-(1/2)^3} = \frac{1}{7}$.

Bài tập 1.11. A, B độc lập nên $P(A \cap B) = P(A).P(B)$. Ta có $P(A) = P((A \cap \bar{B}) \cup (A \cap B)) = P(A \cap \bar{B}) + P(A \cap B) = P(A \cap \bar{B}) + P(A).P(B)$. Suy ra $P(A \cap \bar{B}) = P(A).(1 - P(B)) = P(A).P(\bar{B})$. Vậy A và $\bar{B}$ độc lập.

Bài tập 1.12. Ta lấy 3 sự kiện A, B, C trùng với 3 sự kiện X, Y, Z như trong ví dụ 1.18. Ta có $P(X) = 1/2$, $P(Y) = P(Z) = 1/6$, $P(X \cap Y) = 1/12 = P(X).P(Y), P(X \cap Z) = 1/12 = P(X).P(Z), P(Y \cap Z) = 1/36 = P(Y).P(Z), P(X \cap Y \cap Z) = 0 \neq P(X).P(Y \cap Z)$. Như vậy X độc lập với Y và Z nhưng không độc lập với $Y \cap Z$.

Bài tập 1.13. Gọi C là sự kiện "quân rút ra đầu tiên là quân cơ". Ta có $P(B) = P(B|C).P(C) + P(B|\bar{C}).P(\bar{C}) = \frac{12}{51}.\frac{13}{52} + \frac{13}{51}\frac{39}{52} = \frac{1}{4} < \frac{13}{51} = P(B|A)$. Vậy hai sự kiện A và B không độc lập.

Bài tập 1.14. Gọi A là biến cố người được chọn là đàn ông ($\bar{A}$ là biến cố người được chọn là đàn bà), B là biến cố người được chọn là thừa cân. Theo công thức xác suất toàn phần $P(B) = P(B|A).P(A) + P(B|\bar{A}).P(\bar{A}) = 0.65 \times 0.5 + 0.534 \times 0.5 = 0.592$.

Bài tập 1.15. Gọi A là biến cố người được chọn là đàn ông, B là biến

cố người được chọn bị mù màu. Ta có $P(B|A) = 0.05, P(B|\bar{A}) = 0.0025$. Xác suất để một người mù màu được chọn là đàn ông là $P(A|B) = \dfrac{P(B|A).P(A)}{P(B|A).P(A) + P(B|\bar{A}).P(\bar{A})} = 0.9524.$

Bài tập 1.16. Theo định lý 1.4, $P(B^{\epsilon}_{n,k}) \to 1$ khi $n \to \infty$ $(k = 1, 2, ..., s)$. Do đó $P(B^{\bar{\epsilon}}_{n,k}) \to 0$ khi $n \to \infty$ $(k = 1, 2, ..., s)$. Suy ra $P(\Omega \backslash (B^{\epsilon}_{n,1} \cap B^{\epsilon}_{n,2} \cap ... \cap B^{\epsilon}_{n,s})) = P(B^{\bar{\epsilon}}_{n,1} \cup B^{\bar{\epsilon}}_{n,2} \cup ... \cup B^{\bar{\epsilon}}_{n,k}) \leq P(B^{\bar{\epsilon}}_{n,1}) + P(B^{\bar{\epsilon}}_{n,2}) + ... + P(B^{\bar{\epsilon}}_{n,k}) \to 0$ khi $n \to \infty$. Vậy $P(B^{\epsilon}_{n,1} \cap B^{\epsilon}_{n,2} \cap ... \cap B^{\epsilon}_{n,s}) \to 1$ khi $n \to \infty$.

Bài tập 1.17. Không gian mẫu Ω gồm một dãy kết quả những lần tung, trong đó lần tung cuối cùng thu được mặt ngửa, trong những lần tung trước, có hai lần mặt ngửa xuất hiện.

$$\Omega = \{NNN, SNNN, NSNN, NNSN, SSNNN, ...\}.$$

Để tung sáu lần thì trong năm lần đầu tiên có hai lần mặt ngửa xuất hiện, lần cuối cùng thu được mặt ngửa. Vậy

$$P(A) = C_5^2.\left(\frac{1}{2}\right)^2.\left(\frac{1}{2}\right)^3.\frac{1}{2} = 0.15625.$$

Bài tập 1.18. Gọi Ω là tập hợp những người mua bảo hiểm trong đó, A là tập hợp những người trẻ, B là tập hợp đàn ông, C là tập hợp những người đã có vợ hoặc chồng. Ta có $|\Omega| = 20000, |A| = 6300, |B| = 9600, |C| = 13800, |A \cap B| = 2700, |B \cap C| = 6400, |A \cap C| = 2900, |A \cap B \cap C| = 1100$. Xác suất để một người mua bảo hiểm của hãng là phụ nữ trẻ độc thân là $P(A \cap \bar{B} \cap \bar{C}) = P(A \backslash ((A \cap B) \cup (A \cap C))) = P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C) = \dfrac{6300 - 2700 - 2900 + 1100}{20000} = 0.09.$

Bài tập 1.9. Có thể giải thích rằng sau khi có xe bus đến nhà cô B 7,5 phút thì có xe bus đến nhà cô A.

Bài tập 1.20. i) Xác suất để trong 100 lần quay không có lần nào số 68 trúng giải là $(\frac{99}{100})^{100} \approx 0.366$.
ii) Dành cho bạn đọc tự làm.

Bài tập 1.21. Gọi Ω là không gian mẫu, A_i là biến cố người thứ i nhặt được mũ của mình ($i = 1, 2, ..., n$). Ta cần tính $P(\Omega \backslash (A_1 \cup A_2 \cup ... \cup A_n)) = 1 - P(A_1 \cup A_2 \cup ... \cup A_n)$.
Ta có $P(A_1 \cup A_2 \cup ... \cup A_n)$
$= \Sigma_{i=1}^n P(A_i) - \Sigma_{i,j=1,i \neq j}^n P(A_i \cap A_j) + ... + (-1)^{n-1} P(A_1 \cap A_2 \cap ... \cap A_n)$
$= \frac{1}{n!}(C_n^1.(n-1)! - C_n^2.(n-2)! + ... + (-1)^{n-1}.C_n^n.0!)$
$= 1 - \frac{1}{2!} + ... + \frac{(-1)^{n-1}}{n!}$.
Vậy xác suất để không có người nào nhặt được mũ của chính mình là $1 - (1 - \frac{1}{2!} + ... + \frac{(-1)^{n-1}}{n!}) = \frac{1}{2!} - \frac{1}{3!} + ... + \frac{(-1)^n}{n!} \to e^{-1}$ khi $n \to \infty$.

Bài tập 1.22. Đặt $B_n = \bigcup_{m=n}^{\infty} A_m$, $n = 1, 2, ...$ thì $B_\infty = \bigcap_{n=1}^{\infty} B_n$. Mặt khác $B_1 \supseteq B_2 \supseteq B_3 \supseteq ...$ nên $P(B_\infty) = \lim_{n \to \infty} P(B_n)$.

i) Vì $\Sigma_{n=1}^{\infty} P(A_n) < \infty$ nên $P(B_n) < \Sigma_{m=n}^{\infty} P(A_m) \to 0$ khi $n \to \infty$. Vậy $P(B_\infty) = 0$.

ii) Nếu tồn tại một dãy vô hạn các A_n sao cho $P(A_n) \geq \epsilon$ thì $P(B_n) \geq \epsilon$ với mọi n. Vậy $P(B_\infty) \geq \epsilon$.

Bài tập 1.23. Gọi A là sự kiện ngăn kéo được rút ra là ngăn kéo chứa hai đồng tiền vàng, B là sự kiện đồng tiền rút ra là đồng tiền vàng. Ta cần tính $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/3}{1/2} = \frac{2}{3}$.

Bài tập 1.24. Cai ngục lý luận như vậy là sai. Bài toán này tương tự

như bài toán chơi mở cửa có quà ở đầu Chương 1. Trong hai người B và C luôn có ít nhất 1 người được thả, và nếu nói tên 1 người được thả trong hai người B và C ra, thì xác suất được thả của người còn lại trong hai người đó giảm xuống thành 1/3 trong khi xác suất để A được thả vẫn giữ nguyên là 2/3.

Bài tập 1.25. Gọi A là biến cố một người trong đám đông là kẻ trộm, B là biến cố một người trong đám đông bị máy nghi là có tội. Ta có $P(A) = \frac{2}{60} = \frac{1}{30}, P(B|A) = 0.85, P(B|\bar{A}) = 0.07$. Ta cần tính $P(A|B)$. Theo công thức Bayes, ta có

$$P(A|B) = \frac{P(B|A).P(A)}{P(B|A).P(A) + P(B|\bar{A}).P(\bar{A})}$$
$$= \frac{0.85 \times 1/30}{0.85 \times 1/30 + 0.07 \times 29/30} \approx 0.295.$$

Bài tập 1.26. Gọi X là biến cố một con bò bị mắc bệnh bò điên, Y là biến cố một con bò phản ứng dương tính với xét nghiệm A. Ta có $P(Y|X) = 0.7, P(Y|\bar{X}) = 0.1, P(X) = 1.3 \times 10^{-5}$. Ta cần tính $P(X|Y) = \frac{P(Y|X).P(X)}{P(Y|X).P(X) + P(Y|\bar{X}).P(\bar{X})}$. Kết quả là:

$$P(X|Y) = \frac{0.7 \times 1.3 \times 10^{-5}}{0.7 \times 1.3 \times 10^{-5} + 0.1 \times (1 - 1.3. \times 10^{-5})} \approx 0.000091.$$

Bài tập 1.27. Giả sử $\{Gx, x = 0, 1, 2, ..., 9\}$ là một họ các sự kiện độc lập. Vì giá dầu không thể tăng ít nhất 50% mỗi năm trong 10 năm liên tiếp (như thế giá dầu sẽ vượt quá $10 \times (1.5)^{10} > 300$ USD một thùng) nên $P(G0 \cap G1 \cap ... \cap G9) = P(G0).P(G1) \dots P(G9) = 0$. Suy ra tồn tại $x \in \{0, 1, 2, \dots, 9\}$ để $P(Gx) = 0$. Như vậy nếu coi giá dầu biến động một cách ngẫu nhiên và việc giá dầu tăng ít nhất 50%

trong một năm là hoàn toàn có thể xảy ra (xác suất lớn hơn 0) thì họ các sự kiện trên là không độc lập.

1.2 Lời giải bài tập Chương 2

Bài tập 2.2. Hàm mật độ của X:

$$\rho_X(x) = \begin{cases} 0 \text{ nếu } |x| > 1 \\ 1 - |x| \text{ nếu } |x| \leq 1 \end{cases}$$

Hàm phân phối xác suất của X:

$$P(X \leq x) = \begin{cases} 0 \text{ nếu } x < -1 \\ (1+x)^2/2 \text{ nếu } -1 \leq x \leq 0 \\ 1 - (1-x)^2/2 \text{ nếu } 0 < x < 1 \\ 1 \text{ nếu } x \geq 1 \end{cases}$$

Biến ngẫu nhiên $Y = \arcsin X$ có hàm phân phối:

$$P(Y \leq x) = P(X \leq \sin x) \begin{cases} 0 \text{ nếu } x < -\pi/2 \\ (1+\sin x)^2/2 \text{ nếu } -\pi/2 \leq x \leq 0 \\ 1 - (1-\sin x)^2/2 \text{ nếu } 0 < x < \pi/2 \\ 1 \text{ nếu } x \geq \pi/2 \end{cases}$$

Hàm mật độ của Y là:

$$\rho_Y(x) = \begin{cases} \sin x \cos x + \cos x \text{ nếu } x \in [-\frac{\pi}{2}, 0] \\ -\sin x \cos x + \cos x \text{ nếu } x \in [0, \frac{\pi}{2}] \\ 0 \text{ nếu } |x| > \pi/2 \end{cases}$$

Bài tập 2.3. Nếu phân bố xác suất của biến ngẫu nhiên X là đối xứng và liên tục thì ta có

$$P(X \leq x) = P(X < x) = P(-X < x) = P(X > -x) = 1 - P(X \leq -x),$$

do đó $\mathcal{F}_X(x) + \mathcal{F}_X(-x) = 1$.

Nếu F không liên tục, khi đó kết luận trên không còn đúng. Phản ví dụ: X có phân bố xác suất tập trung tại $x = 0$, tức $P(X = 0) = 1$, khi đó: $F(x) + F(-x) = 2$ với $x = 0$.

Bài tập 2.5. Với mỗi $y \in [0, 1]$, đặt $g(y) = \sup\{x : F_Y(x) < y\}$. Ta chứng minh: $g(y) \leq z \Leftrightarrow F_Y(z) \geq y$. Thật vậy :

- Giả sử $F_Y(z) < y$. Do F liên tục phải nên $\exists z' > z$ sao cho $F_Y(z) < F_Y(z') < y$
$\Rightarrow z < z' \leq \sup\{x : F_Y(x) < y\} = g(y)$.
- Giả sử $F_Y(z) \geq y$. Khi đó $z > x$ với mọi x thỏa mãn $F_Y(x) < y$
$\Rightarrow z \geq \sup\{x : F_Y(x) < y\} = g(y)$.

Ta có $F_{g(X)}(z) = P(g(X) \leq z) = P(F_Y(z) \geq X) = P(X \leq F_Y(z)) = F_Y(z)$
(Vì X có phân phối đều $\mathcal{U}(0, 1)$, $F(x) = x$ với mọi $x \in [0, 1]$.) Hàm g định nghĩa như trên chính là hàm cần tìm.

Bài tập 2.6. $X \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow$ hàm mật độ của X:

$$\rho_X(x) = \frac{1}{2\pi\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

Xét $y = f(x) = \dfrac{x-\mu}{\sigma} \Rightarrow f'(x) = \dfrac{1}{\sigma}$. Ta có: $\rho_Y(y) = \dfrac{\rho_X(x)}{|f'(x)|} = \dfrac{1}{2\pi\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \sigma = \frac{1}{2\pi} e^{-\frac{1}{2}y^2}$. Vậy Y có phân phối chuẩn $\mathcal{N}(0, 1)$.

Bài tập 2.7. $X \sim \mathcal{E}(\lambda) \Rightarrow$ hàm mật độ:

$$\rho_X(x) = \begin{cases} \lambda e^{-\lambda x} \text{ nếu } x > 0 \\ 0 \text{ nếu } x \leq 0 \end{cases}$$

$Y = cX, c > 0$. Xét $y = f(x) = cx$.

$$\rho_Y(y) = \begin{cases} \dfrac{\lambda e^{-\lambda x}}{c} \text{ nếu } x > 0 \\ 0 \text{ nếu } x \leq 0 \end{cases} = \begin{cases} \dfrac{\lambda}{c} e^{-\frac{\lambda}{c} \cdot y} \text{ nếu } y > 0 \\ 0 \text{ nếu } y \leq 0 \end{cases}$$

Vậy $Y \sim \mathcal{E}(\frac{\lambda}{c})$.

Bài tập 2.8. Ta có:

$$\begin{aligned} P(X > s + t | X > s) &= \frac{P(X > s + t, X > s)}{P(X > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t) \end{aligned}$$

Bài tập 2.9. i) Tương tự như bài tập 2.5.

ii) Nếu $X \sim \mathcal{U}(0, 1)$, tức là có hàm mật độ $\rho_X(x) = 1$ trên đoạn thẳng $[0, 1]$, và $Y = -\ln X \sim \mathcal{E}(1)$ thì hàm mật độ của Y là:

$$\begin{aligned} \rho_Y(y) &= \frac{\rho_X(x)}{|\frac{d \ln x}{dx}|} = \begin{cases} \dfrac{1}{1/x} = x \text{ nếu } x \in [0, 1] \\ 0 \text{ nếu } x \notin [0, 1] \end{cases} \\ &= \begin{cases} e^{-y} \ \forall y > 0 \\ 0 \ \forall y \leq 0 \end{cases} \end{aligned}$$

Điều đó có nghĩa là Y có phân bố xác suất mũ $\mathcal{E}(1)$.

Bài tập 2.10. X có phân phối Pareto với tham số α và hàm mật độ là

$$\rho_X(x) = \begin{cases} \dfrac{\alpha}{x^{\alpha+1}} \text{ nếu } x \geq 1 \\ 0 \text{ nếu } x < 1 \end{cases}$$

$Y = X^s, s > 0$. Xét $y = f(x) = x^s$, ta có:

$$\begin{aligned} \rho_Y(y) &= \frac{\rho_X(x)}{|f'(x)|} = \begin{cases} \dfrac{\frac{\alpha}{x^{\alpha+1}}}{s \cdot x^{s-1}} = \dfrac{\alpha}{s} \cdot \dfrac{1}{x^{\alpha+s}} \text{ nếu } x \geq 1 \\ 0 \text{ nếu } x < 1 \end{cases} \\ &= \begin{cases} \dfrac{\alpha}{s} \cdot \dfrac{1}{y^{\frac{\alpha}{s}+1}} \text{ nếu } y \geq 1 \\ 0 \text{ nếu } y < 1 \end{cases} \end{aligned}$$

(Do $x = y^{\frac{1}{s}} > 1 \Leftrightarrow y > 1$). Vậy Y có phân phối Pareto với tham số $\dfrac{\alpha}{s}$.

Bài tập 2.11. $X \sim \mathcal{U}(0,1), Y = \dfrac{1}{1-X}$. Xét $y = f(x) = \dfrac{1}{1-x} \Rightarrow f'(x) = \dfrac{1}{(1-x)^2}$

Ta có

$$\rho_Y(y) = \frac{\rho_X(x)}{|f'(x)|} = \begin{cases} (1-x)^2 = \frac{1}{y^2} \text{ nếu } y \geq 1 \\ 0 \text{ nếu } y < 1 \end{cases}$$

$\Rightarrow Y$ có phân phối Pareto với tham số $\alpha = 1$.

Bài tập 2.12. Kì vọng lợi nhuận:

$$E = 0.7 \times (0 - 100000) + 0.3 \times (1000000 - 100000) = 200000.$$

Bài tập 2.13. Điểm cần chú ý khi lấy ví dụ là, cần cho phân bố xác suất chung của X và Y chứ không cho phân bố xác suất của riêng X

và riêng Y rồi coi hai biến đó độc lập với nhau. (Nếu chúng độc lập thì $\mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY)$). Chẳng hạn có thể chọn X và Y là những biến ngẫu nhiên nhận hai giá trị 0 và 1, với phân bố xác suất chung như sau: $P(X = 0, Y = 0) = 0.2, P(X = 0, Y = 1) = 0.4, P(X = 1, Y = 0) = 0.3, P(X = 1, Y = 1) = 0.1$ Khi đó $\mathbb{E}(X) = 0.4, \mathbb{E}(Y) = 0.5, \mathbb{E}(XY) = 0.1 \neq \mathbb{E}(X)\mathbb{E}(Y) = 0.2$.

Bài tập 2.14. Có 99 quả được đánh số từ 1 đến 99, lấy ngẫu nhiên 5 quả, ta có lực lượng của không gian mẫu là $|\Omega| = C_{99}^5$. Gọi hai biến ngẫu nhiên: X là "số nhỏ nhất trên 5 quả bốc được", Y là "số lớn nhất trên 5 quả bốc được".

i) Phân bố xác suất của X và Y:

X	1	2	$\cdots$	95
p	$\frac{C_{98}^4}{C_{99}^5}$	$\frac{C_{97}^4}{C_{99}^5}$	$\cdots$	$\frac{C_4^4}{C_{99}^5}$

Y	5	6	$\cdots$	99
p	$\frac{C_4^4}{C_{99}^5}$	$\frac{C_5^4}{C_{99}^5}$	$\cdots$	$\frac{C_{98}^4}{C_{99}^5}$

Ví dụ, nếu $X = 2$ thì có nghĩa là có 1 quả trong 5 quả bóng bốc ra là số 2, còn 4 quả còn lại nằm trong các số từ 3 đến 99. Có C_{97}^4 cách chọn 4 số khác nhau trong 97 số từ 3 đến 99, có nghĩa là tập hợp các khả năng với $X = 2$ có C_{97}^4 phần tử trong không gian xác suất có C_{97}^5 phần tử với phân bố xác suất đều, do đó ta có $P(X = 2) = C_{97}^4/C_{99}^5$.

ii) Công thức $\sum_{k=m}^n C_k^m = C_{n+1}^{m+1}$ sinh ra từ công thức $C_{k+1}^{m+1} = C_k^{m+1} + C_k^m$.

iii) Sử dụng công thức trên, ta có:

$$\begin{aligned}\mathbb{E}(X) &= \frac{C_{98}^4 + 2C_{97}^4 + \cdots + 95C_4^4}{C_{99}^5} \\ &= \frac{\sum\limits_{k=4}^{98} C_k^4 + \sum\limits_{k=4}^{97} C_k^4 + \cdots C_4^4}{C_{99}^5} \\ &= \frac{C_{99}^5 + C_{98}^5 + \cdots + C_5^5}{C_{99}^5} \\ &= \frac{C_{100}^6}{C_{99}^5} = \frac{100!}{6! \cdot 94!} \cdot \frac{94! \cdot 5!}{99!} = \frac{100}{6}.\end{aligned}$$

Bài tập 2.15. Phân bố xác suất của X:

$$P(X = k) = C_6^k (\frac{2}{3})^k (\frac{1}{3})^{6-k},$$

$(k = 0, 1, 2, \ldots, 6)$. Vì $Z = X - Y$ và $X + Y = 6$ nên $Z = 2X - 6$, và ta có thể viết phân bố xác suất của Z:

$$P(Z = 2k - 6) = C_6^k (\frac{2}{3})^k (\frac{1}{3})^{6-k}, \quad k = 0, \ldots, 6$$

Từ đó tính được ra kỳ vọng của Z:

$$\mathbb{E}(Z) = \sum_{k=0}^{6} C_6^k (\frac{2}{3})^k (\frac{1}{3})^{6-k} (2k - 6) = 2.$$

Một cách tính đơn giản hơn là: kỳ vọng để bóng vào rổ mỗi lần ném là 2/3. Bởi vậy nếu ném 6 lần, thì kỳ vọng số lần bóng vào rổ là $6 \times 2/3 = 4$, tức là ta có $\mathbb{E}(X) = 4$, từ đó suy ra $\mathbb{E}(Z) = 2\mathbb{E}(X) - 4 = 2$.

Bài tập 2.16. i) Chiến thuật:
Lần thứ nhất: B hỏi: "số đó có lớn hơn 2^{n-1} không?"

- Nếu câu trả lời là "có": số đó sẽ nằm trong đoạn $\left[2^{n-1}+1; 2^n\right]$
- Nếu câu trả lời là "không": số đó sẽ nằm trong đoạn $\left[1; 2^{n-1}\right]$

Lần thứ i: Ta sẽ xác định được số đó nằm trong đoạn có độ dài 2^{n-i}. Vậy, sau n lần, ta sẽ xác định được số A đã chọn.

ii) Ta sẽ chứng minh một khẳng định tổng quát hơn: giả sử X là một tập hữu hạn có m phần tử, A chọn một phần tử của X, và B hỏi các câu hỏi kiểu "phần tử đó có nằm trong tập con Y của X không", và A sẽ trả lời là có hoặc không. Khi đó mọi chiến thuật hỏi của B sẽ cần trung bình ít nhất là $\log_2 m$ câu hỏi để xác định phần tử mà A chọn. (Trường hợp bài toán nêu ra là trường hợp $m = 2^n$).

Ta có thể chứng minh khẳng định này bằng cách qui nạp theo m. Với các số m nhỏ (ví dụ $m = 2$ hay $m = 3$), dễ dàng kiểm tra trực tiếp khẳng định), và với $m = 1$ thì khẳng định là hiển nhiên. Giả sử ta đã chứng minh được khẳng định cho các tập có không quá $m-1$ phần tử ($m \geq 2$), ta sẽ chứng minh rằng khẳng định đúng cho tập X với m phần tử.

Dù là chiến thuật nào, thì câu đầu tiên của B cũng phải có dạng "phần tử đó có nằm trong Y không", trong đó Y là một tập con của X mà B chọn ra. Nếu câu trả lời là có, thì trong các bước tiếp theo B phải chọn các tập con của Y, và như vậy, theo qui nạp, sẽ cần thêm trung bình ít nhất là $\log_2 |Y|$ lần hỏi. Ở đây ta có thể coi rằng $1 \leq |Y| = k < |X| = m$. Nếu câu trả lời là không, thì có nghĩa là phần từ A chọn nằm trong $X \setminus Y$, và sẽ cần thêm trung bình ít nhất $\log_2 |X \setminus Y| = \log_2(m-k)$ lần hỏi. Xác suất để phần tử mà A chọn rơi vào Y, tức là để A trả lời Yes cho câu hỏi đầu tiên là $|Y|/|X| = k/m$,

và xác suất để A trả lời No cho câu hỏi đầu tiên là $(m-k)/m$. Như vậy, nếu trong chiến thuật hỏi dùng tập con Y cho câu hỏi đầu tiên, thì sẽ cần trung bình ít nhất là

$$1 + \frac{k}{m}\log_2 k + \frac{m-k}{m}\log_2(m-k)$$

câu hỏi để xác định được phần tử A chọn. Chú ý rằng hàm $x\log_2 x$ là hàm lồi, do đó khi m cố định và $0 < k < m$ thì giá trị của $\frac{k}{m}\log_2 k + \frac{m-k}{m}\log_2(m-k)$ đạt cực tiểu khi mà $k = m-k = m/2$, bởi vậy ta có

$$1 + \frac{k}{m}\log_2 k + \frac{m-k}{m}\log_2(m-k) \geq 1 + 2.(1/2).\log_2(m/2) = \log_2 m,$$

từ đó suy ra điều phải chứng minh.

Bài tập 2.17.

$$\rho_Y(x) = \begin{cases} c\sin x \text{ nếu } x \in (0,\pi) \\ 0 \text{ nếu } x \notin (0,\pi) \end{cases}$$

i) Ta có:

$$1 = \int_{-\infty}^{\infty} \rho_Y(x)dx = \int_0^{\pi} c\sin x dx = -c\cos x\Big|_0^{\pi} = 2c,$$

do đó $c = 1/2$.

ii) $\mathbb{E}(Y) = \int_0^{\pi} \frac{1}{2}x\sin x dx = \frac{\pi}{2}$. Một các giải thích khác là, hàm mật độ $\rho_Y(x)$ đối xứng quanh điểm $x = \pi/2$, và do đó giá trị kỳ vọng của Y bằng $\pi/2$.

Bài tập 2.18. X có phân phối Pareto với tham số $\alpha > 1$, hàm mật độ:

$$\rho_X(x) = \begin{cases} \frac{\alpha}{x^{\alpha+1}} \text{ nếu } x \geq 1 \\ 0 \text{ nếu } x < 1 \end{cases}$$

Kỳ vọng của X là:

$$E(X) = \int_1^\infty x \frac{\alpha}{x^{\alpha+1}} dx = \frac{\alpha}{-\alpha+1} x^{-\alpha+1}\Big|_1^\infty = \frac{\alpha}{\alpha-1}.$$

Bài tập 2.19. Gọi (a_i, b_i), $i = 1, \ldots, n$, là các cặp giá trị của (F, G), $a_i, b_i > 0$. Từ giả thiết ta có $P(F = a_i, G_i = b_i) = \frac{1}{n}$, và

$$\mathbb{G}(F) = \sqrt[n]{\prod_{i=1}^n a_i}, \mathbb{G}(G) = \sqrt[n]{\prod_{i=1}^n b_i}, \mathbb{G}((F+G)/2) = \sqrt[n]{\prod_{i=1}^n \frac{a_i + b_i}{2}}.$$

Áp dụng bất đẳng thức Cauchy (trung bình nhân nhỏ hơn trung bình cộng), ta có

$$\begin{aligned}\frac{\mathbb{G}(F) + \mathbb{G}(G)}{2\mathbb{G}((F+G)/2)} &= \sqrt[n]{\prod_{i=1}^n \frac{a_i}{a_i + b_i}} + \sqrt[n]{\prod_{i=1}^n \frac{b_i}{a_i + b_i}} \\ &\leq \frac{1}{n}\left(\sum_{i=1}^n \frac{a_i}{a_i + b_i}\right) + \frac{1}{n}\left(\sum_{i=1}^n \frac{b_i}{a_i + b_i}\right) = 1.\end{aligned}$$

Bài tập 2.20. i) X có phân phối hình học $P(k) = p(1-p)^{(k-1)}$.

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k\geq 1} k \cdot p(1-p)^{k-1} = p\sum_{k\geq 1} k \cdot (1-p)^{k-1} \\ &= p\frac{1}{(1-(1-p))^2} = \frac{1}{p}, \\ \mathbb{E}(X^2) &= \sum_{k\geq 1} k^2 \cdot p(1-p)^{k-1} \\ &= p(1-p)\sum_{k\geq 2} k(k-1) \cdot (1-p)^{k-2} + p\sum_{k\geq 1} k \cdot p(1-p)^{k-1} \\ &= p(1-p)\frac{2}{(1-(1-p))^3} + \frac{1}{p} = \frac{2-p}{p^2}, \\ var(X) &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{1-p}{p^2}.\end{aligned}$$

Do đó, độ lệch chuẩn $\sigma = \sqrt{var(X)} = \dfrac{\sqrt{1-p}}{p}$.

ii) X có phân phối Poisson $P(k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$.

$$\begin{aligned}
\mathbb{E}(X) &= e^{-\lambda}\sum_{k\geq 1} k\cdot\frac{\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{k\geq 1}\frac{\lambda^{k-1}}{(k-1)!}\\
&= \lambda e^{-\lambda}e^{\lambda} = \lambda,\\
\mathbb{E}(X^2) &= e^{-\lambda}\sum_{k\geq 1} k^2\cdot\frac{\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{k\geq 1} k\cdot\frac{\lambda^{k-1}}{(k-1)!}\\
&= \lambda e^{-\lambda}\left[\sum_{k\geq 1}(k-1)\cdot\frac{\lambda^{k-1}}{(k-1)!} + \sum_{k\geq 1}\frac{\lambda^{k-1}}{(k-1)!}\right]\\
&= \lambda e^{-\lambda}(\lambda e^{\lambda} + e^{\lambda}) = \lambda^2 + \lambda\\
\Rightarrow \sigma &= \sqrt{\mathbb{E}(X^2) - (\mathbb{E}(X))^2} = \sqrt{\lambda^2 + \lambda - \lambda^2} = \sqrt{\lambda}.
\end{aligned}$$

Bài tập 2.21. $\mathbb{E}(X) = 2/3$, $\rho_X(x) = \begin{cases} ax^2 + b \text{ nếu } 0 < x < 1 \\ 0 \text{ nếu không} \end{cases}$

Ta có

$$2/3 = \mathbb{E}(X) = \int\limits_0^1 x(ax^2+b)dx = \frac{a}{4} + \frac{b}{2}$$

và

$$1 = \int\limits_{-\infty}^{\infty} \rho_X(x)dx = \int\limits_0^1 (ax^2+b)dx = \frac{a}{3} + b.$$

Giải hệ phương trình tuyến tính trên theo a và b, ta được $a = 2, b =$

$1/3$, từ đó suy ra

$$\mathbb{E}(X^2) = \int_0^1 x^2(2x^2 + 1/3)dx = 2/5 + 1/9 = 23/45$$

và $var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 23/45 - 4/9 = 1/15$.

Bài tập 2.22. i) Xét một mẫu máu hỗn hợp gồm k mẫu máu. Ta gọi các biến cố: A = "mẫu máu hỗn hợp chứa kháng thể X"; A_i = "mẫu máu i chứa X". Khi đó

$$P(\bar{A}) = \prod_{i=1}^{k} P(\bar{A}_i) = (1-p)^k.$$

Để cho gọn, đặt $(1-p)^k = q$, ta có $P(A) = 1 - (1-p)^k = 1 - q$.

ii) Gọi S là biến ngẫu nhiên tổng số lần phải xét nghiệm. Ta có phân phối của S:

S	m	$m+k$	$\cdots$	$m+mk$
p	q^m	$C_m^1 q^{m-1}(1-q)$	$\cdots$	$(1-q)^m$

$$\Rightarrow \mathbb{E}(S) = \sum_{i=0}^{m} C_m^i q^{m-i}(1-q)^i(m+ik)$$
$$= m\sum_{i=0}^{m} C_m^i q^{m-i}(1-q) + k\sum_{i=0}^{m} iC_m^i q^{m-i}(1-q)^i$$
$$= m + k(1-q)m\sum_{i=1}^{m} C_{m-1}^{i-1} q^{m-i}(1-q)^{i-1}$$
$$= m + mk(1-q),$$
$$\mathbb{E}(S^2) = \sum_{i=0}^{m} C_m^i q^{m-i}(1-q)^i(m+ik)^2$$
$$= m^2 + 2mk\sum_{i=0}^{m} iC_m^i q^{m-i}(1-q)^i + k^2\sum_{i=0}^{m} i^2C_m^i q^{m-i}(1-q)^i$$
$$= m^2 + 2mk\cdot m(1-q) + k^2\left[m(1-m)(1-q)^2 + m(1-q)\right]$$
$$= m^2 + 2m^2k(1-q) + mk^2(1-q)\left[(m-1)(q-1)+1\right]$$
$$\Rightarrow var(S) = \mathbb{E}(S^2) - \mathbb{E}(S)^2 = mk^2(1-q)q.$$

iii) Ta có $\mathbb{E}(S) < N \Leftrightarrow m+mk(1-q) < mk \Leftrightarrow kq > 1 \Leftrightarrow k(1-p)^k > 1$.

Bài tập 2.23. Không mất tính tổng quát, ta coi X, Y nhận hữu hạn các giá trị $a_1, a_2, \ldots, a_n$, $(a_i \neq a_j \ \forall\ i \neq j)$, với các xác suất $p(X = a_i) = p_i \geq 0$ và $p(Y = a_i) = q_i \geq 0$. Ở đó $\sum\limits_{i=1}^{n} p_i = \sum\limits_{i=1}^{n} q_i = 1$. Theo giả thuyết, ta có $E(X^k) = \sum_i a_i^k p_i = E(Y^k) = \sum_i a_i^k q_i$, hay có nghĩa là $\sum_{i=1}^{n} a_i^k x_i = 0$ với $x_i = p_i - q_i$ và với mọi số tự nhiên k. Cho k chạy từ 1 đến n, ta được một hệ phương trình tuyến tính với n ẩn x_i và n phương trình.

Nếu $a_i \neq 0$ với mọi i, thì định thức của hệ phương trình này là

định thức của ma trận

$$\begin{vmatrix} a_1 & a_2 & \cdots & a_n \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \cdots & \cdots & \cdots & \cdots \\ a_1^n & a_2^n & \cdots & a_n^n \end{vmatrix}$$

(gọi là định thức Vandermonde) có giá trị khác 0 vì các số a_i khác nhau, và do đó hệ phương trình chỉ có một nghiệm duy nhất là nghiệm tầm thường $x_i = 0$ với mọi i, có nghĩa là ta có $p_i = q_i$ với mọi i, hay nói cách khác, X và Y có cùng phân bố xác suất.

Nếu giả sử chẳng hạn $a_1 = 0$, thì ta chỉ xét $n-1$ phương trình đầu tiên, với $n-1$ ẩn số $x_2, \ldots, x_n$. Tương tự như trường hợp phía trên, ta phải có $p_i = q_i$ với mọi $i \geq 2$, từ đó suy ra $p_1 = 1 - \sum_{i\geq 2} p_i = 1 - \sum_{i \geq 2} q_i = q_1$, và X và Y cũng có cùng phân bố xác suất.

Bài tập 2.24. Giả sử X có phân bố mũ với tham số $\lambda > 0$, hàm mật độ của X là

$$\rho_X(x) = \begin{cases} \lambda e^{-\lambda x} \text{ nếu } x > 0 \\ 0 \text{ nếu } x \leq 0 \end{cases}$$

Khi đó $\mathbb{E}(X) = 1/\lambda$, và moment bậc n của X bằng $\mathbb{E}(X) = \lambda^n n!$ với mọi $n \in \mathbb{N}$.

Bài tập 2.25. Hàm đặc trưng của X là $\phi_X(s) = \mathbb{E}(\cos sX) + i\mathbb{E}(\sin sX)$, ($s \in \mathbb{R}$), và của $-X$ là $\phi_{-X}(s) = \mathbb{E}(\cos s(-X)) + i\mathbb{E}(\sin s(-X)) = \overline{\phi_X(s)}$. X đối xứng khi và chỉ khi X và $-X$ có cùng phân bố xác suất, tức là khi và chỉ khi X và $-X$ có cùng hàm đặc trưng, có nghĩa là $\phi_X = \overline{\phi_X}$, hay nói cách khác ϕ_X là hàm thực.

Bài tập 2.26. X có phân phối hình học $P(X = k) = p(1-p)^{k-1} \; \forall\, k \geq$

1. Hàm sinh xác suất của X:

$$G(z) = \sum_{k=1}^{\infty} P(k) z^k = \sum_{k=1}^{\infty} p(p-1)^{k-1} z^k = pz \sum_{k=0}^{\infty} (z-pz)^k = \frac{pz}{1-z+pz}.$$

Từ đó ta có:

$$G'(z) = \frac{p}{(1-z+pz)^2}, G"(z) = \frac{2p(1-p)}{(1-z+pz)^3},$$

và phương sai của X là:

$$var(X) = G"(1) + G'(1) - (G'(1))^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Bài tập 2.27. X có phân phối nhị thức tham số n, p. $P(k) = C_n^k p^k (1-p)^k$. Hàm sinh xác suất của X:

$$G(z) = \sum_{k=0}^{n} z^k P(k) = \sum_{k=0}^{n} z^k C_n^k p^k (1-p)^k = (pz + 1 - p)^n.$$

Hàm Laplace:

$$L(t) = E(e^{-tX}) = \sum_{k=0}^{n} e^{-tk} C_n^k p^k (1-p)^k = (e^{-t} p + 1 - p)^n.$$

1.3 Lời giải bài tập Chương 3

Bài tập 3.1. Ta có:

$$\begin{aligned} &P_F(]a,b] \times]c,d] \times]e,f]) = \\ &\qquad = P_F(]-\infty,b] \times]c,d] \times]e,f]) - P_F(]-\infty,a] \times]c,d] \times]e,f]) \end{aligned}$$

Số hạng thứ nhất:

$$P_F(]-\infty, b]\times]c, d]\times]e, f])$$
$$= P_F(]-\infty, b]\times]-\infty, d]\times]e, f]) - P_F(]-\infty, b]\times]-\infty, c]\times]e, f])$$
$$= P_F(]-\infty, b]\times]-\infty, d]\times]-\infty, f]) - P_F(]-\infty, b]\times]-\infty, d]\times]-\infty, e])$$
$$+ P_F(]-\infty, b]\times]-\infty, c]\times]-\infty, e]) - P_F(]-\infty, b]\times]-\infty, c]\times]-\infty, f]$$
$$= \mathcal{F}_F(b, d, f) - \mathcal{F}_F(b, d, e) + \mathcal{F}_F(b, c, e) - \mathcal{F}_F(b, c, f)$$

Số hạng thứ hai:

$$P_F(]-\infty, a]\times]c, d]\times]e, f])$$
$$= \mathcal{F}_F(a, d, f) - \mathcal{F}_F(a, d, e) + \mathcal{F}_F(a, c, e) - \mathcal{F}_F(a, c, f)$$

Vậy:

$$P_F(]a, b]\times]c, d]\times]e, f])$$
$$= \mathcal{F}_F(b, d, f) - \mathcal{F}_F(b, d, e) + \mathcal{F}_F(b, c, e) - \mathcal{F}_F(b, c, f)$$
$$- \mathcal{F}_F(a, d, f) + \mathcal{F}_F(a, d, e) - \mathcal{F}_F(a, c, e) + \mathcal{F}_F(a, c, f)$$

Bài tập 3.2. Giả sử hai người hẹn gặp nhau tại A.
X là thời điểm người 1 đến A ($\in [0, 1]$)
Y là thời điểm người 2 đến A ($\in [0, 1]$)

12h	13h
0	1

$$\rho_X(x) = \begin{cases} 1, & x \in [0, 1] \\ 0, & x \notin [0, 1] \end{cases}, \quad \rho_Y(y) = \begin{cases} 1, & y \in [0, 1] \\ 0, & y \notin [0, 1] \end{cases}$$

Ta có:

$$P(-\frac{1}{4} \leq X - Y \leq \frac{1}{4}) = P_{X,Y}(-\frac{1}{4} \leq x - y \leq \frac{1}{4})$$
$$= \iint\limits_{-\frac{1}{4} \leq x-y \leq \frac{1}{4}} \rho_{X,Y}(x,y)dxdy = \iint\limits_{\substack{-\frac{1}{4} \leq x-y \leq \frac{1}{4} \\ 0 \leq x \leq 1,\ 0 \leq y \leq 1}} dxdy = \frac{7}{16}$$

Vậy xác suất để hai người gặp nhau theo hẹn là $\frac{7}{16}$.

Bài tập 3.3. Giả sử tồn tại hàm mật độ của (X, X^3) là $\rho_{X,Y}(x,y)$ với $X^3 = Y$. Xét hàm số $f : R^2 \longrightarrow R^2$, $(x,y) \longmapsto (x^3 - y, x + y)$ là song ánh khả vi liên tục, có Jacobian

$$J = \begin{vmatrix} 3x^2 & -1 \\ 1 & 1 \end{vmatrix} = 3x^2 + 1 \neq 0$$

suy ra véc tơ ngẫu nhiên $(U = X^3 - Y, V = X + Y)$ có hàm mật độ

$$\rho_{U,V}(u,v) = \rho_{X,Y}(f^{-1}(u,v)).|J(f^{-1}(u,v))|^{-1}$$

Vậy tồn tại hàm mật độ biên của U là $\rho_U(u) = \int_{-\infty}^{+\infty} \rho_{U,V}(u,v)dv$.
Nhưng $U = 0$ nên là biến ngẫu nhiên rời rạc có điểm hạt là 0 mâu thuẫn. Vậy không tồn tại hàm mật độ đồng thời của X và X^3.
* Thay X^3 bởi $\phi(X)$ với ϕ đơn điệu cũng được kết quả tương tự.

Bài tập 3.4. Tung một xúc sắc 2 lần, được hai số ký hiệu là a, b. Xét ba sự kiện: A là "$a + b$ là số chẵn", B là "$a = 1$", C là "b=4".
Dễ kiểm tra được A, B, C độc lập từng đôi, A và $B \cup C$ không độc lập.
Xét $X = \Psi_A, Y = \Psi_B, X = \Psi_C$, khi đó $Z + Y = \Psi_{B \cup C}$.
Ta có: X độc lập với Y và Z nhưng không độc lập với $Y + Z$.

Bài tập 3.5. Phân bố xác suất đồng thời của 3 biến X, Y, Z độc lập với phân bố xác suất đều trên đoạn $[0,1]$:

$$\rho(x,y,z) = \begin{cases} 1, & (x,y,x) \in [0,1]^3 \\ 0, & (x,y,z) \notin [0,1]^3 \end{cases} .$$

Dễ thấy xác suất cần tìm bằng $P = 1 - P(X+Y \leq Z) - P(X+Z \leq Y) - P(Y+Z \leq X) = 1 - 3P(X+Y \leq Z)$, mặt khác $P(X+Y \leq Z)$ bằng thể tích của hình tứ diện cho bởi các mặt $x \geq 0, y \geq 0, z \geq x+y, 1 \geq z$ trong không gian Euclide $\mathbb{R}^3$ với hệ tọa độ chuẩn (x,y,z), và nó bằng 1/6. Từ đó suy ra $P = 1/2$.

Bài tập 3.6. Có thể chứng minh bằng qui nạp theo k.

Bài tập 3.7. Ta có:

$$\rho_X(x) = \int_{-\infty}^{+\infty} \rho(x,y)dy = \begin{cases} x.e^{-x} & \text{với} \quad x > 0 \\ 0 & \text{với} \quad x \leq 0 \end{cases} ,$$

$$\rho_Y(y) = \int_{-\infty}^{+\infty} \rho(x,y)dx = \begin{cases} e^{-y} & \text{với} \quad y > 0 \\ 0 & \text{với} \quad y \leq 0 \end{cases} .$$

Suy ra $\rho_X(x).\rho_Y(y) = \rho(x,y)$ nên X, Y là hai biến ngẫu nhiên độc lập.

Có thể chứng minh ngắn gọn hơn như sau: hàm mật độ đồng thời $\rho(x,y)$ có thể viết dưới dạng tích của hai hàm, một hàm chỉ phụ thuộc vào x và một hàm chỉ phụ thuộc vào y, do đó X và Y độc lập với nhau.

Bài tập 3.8. Giả sử $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \Rightarrow X_1 = \sigma_1 X + \mu_1$ với $X \sim \mathcal{N}(0,1)$.

Suy ra

$\Phi_{X_1}(s) = \mathbb{E}(\exp(isX_1)) = \mathbb{E}(\exp(is\sigma_1 X + is\mu_1)) = \exp(is\mu_1).\phi_X(\sigma_1 s)$

Vậy $\Phi_{X_1}(s) = \exp(is\mu_1 - \frac{\sigma_1^2 s^2}{2})$, tương tự $\Phi_{X_2}(s) = \exp(is\mu_2 - \frac{\sigma_2^2 s^2}{2})$
Vì X_1, X_2 độc lập nên $\Phi_{X_1+X_2}(s) = \Phi_{X_1}(s).\Phi_{X_2}(s) = \exp(is(\mu_1 + \mu_2) - \frac{(\sigma_1^2+\sigma_1^2)s^2}{2})$ hay $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \frac{\sigma_1^2+\sigma_1^2}{2})$.

Bài tập 3.9. Xác suất cần ước lượng là xác suất để tổng của các số hiện lên trong 350 lần tùng đầu nhỏ hơn 1000.

Bài tập 3.10. Gọi $X_1, ..., X_5$ là các bnn chỉ số chấm xuất hiện trong 5 lần tung. Khi đó $X = X_1 + X_2 + X_3 + X_4 + X_5$ chỉ tổng số chấm xuất hiện trong 5 lần tung.
Ta có:

$$G_{X_i}(z) = \mathbb{E}(z^{X_1}) = \frac{1}{6}\sum_{k=1}^{6} z^k$$

Suy ra:

$$G_X(z) = \prod_{i=1}^{5} G_{X_i}(z) = \frac{1}{6^5}(\sum_{k=1}^{6} z^k)^5$$

Vậy xác suất cần tính là hệ số của z^{15} trong khai triển của $G_X(z)$.

Bài tập 3.11. Đặt $Y_i = X_i - \mu$ thì $\mathbb{E}(Y_i) = 0$, $V(Y) = \sigma^2$. Chọn n đủ lớn để $c - n\mu > 0$. Ta có $P(S_n \geq c) = P(S_n - n\mu \geq c - n\mu) \leq P(|S_n - n\mu| \geq c - n\mu) = P(|\sum_1^n Y_i| \geq c - n\mu) \leq \dfrac{\mathbb{E}(|\sum_1^n Y_i|)^2}{(c - n\mu)^2}$. Mà $\mathbb{E}(|\sum_1^n Y_i|)^2 = n\sigma^2$, vậy $P(S_n \geq c) \leq \dfrac{n\sigma^2}{(c - n\mu)^2} \to 0$ khi $n \to +\infty$

Bài tập 3.12. Gọi X_i là bnn chỉ số tiền thu được ở ngày thứ i của năm.

Ta có:

$$\mu = \mathbb{E}(X_i) = \frac{19}{37}.50 - \frac{18}{37}.50 = \frac{50}{37}$$

$$\sigma^2 = var(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 < \infty$$

Theo luật số lớn, với $n = 365$:

$$\frac{S_n}{n} \approx \mu \Rightarrow S_n \approx \frac{18250}{37} \approx 493$$

là ước lượng số tiền thu về được trong 1 năm (theo đơn vị nghìn euro).

Bài tập 3.13. Ví dụ: Xét biến ngẫu nhiên X có phân phối như sau

X	-1	0	1
p	1/4	1/2	1/4

khi đó phân phối của X^2 là

X	0	1
p	1/2	1/2

Ta có $cov(X, X^2) = E(X^3) - E(X)E(X^2) = 0$ nhưng X, X^2 không độc lập (vì $P(X = -1, X^2 = 1) = P(X = -1) = 1/4 \neq P(X = -1)P(X^2 = 1)$).

Bài tập 3.14. $r(X, Y) \approx -0.9427$.

Bài tập 3.17. Ta có:

$$\begin{aligned} P(K = k) &= \sum_{n \geq k} P(K = k \mid N = n).P(N = n) \\ &= \sum_{n \geq k} C_n^k p^k (1-p)^{n-k}.\frac{e^{-\lambda}\lambda^n}{n!} \\ &= \frac{p^k e^{-\lambda}\lambda^k}{k!}.\sum_{n \geq k} \frac{(q\lambda)^{n-k}}{(n-k)!} = \frac{p^k e^{-\lambda}\lambda^k}{k!}.e^{qk} = \frac{(p\lambda)^k e^{-p\lambda}}{k!}. \end{aligned}$$

Vậy K có phân phối Poisson với tham số $p\lambda$.

Bài tập 3.18. Ta có:
$\mathbb{E}(Y) = \int_R \mathbb{E}(Y \mid X = x)\rho_X(x)dx = \int_{R^+} \frac{x}{2}.\lambda e^{-\lambda x}dx = \frac{1}{2\lambda}$.

Bài tập 3.19. Giả sử $X \sim \mathcal{N}(\mu, \Sigma), Z = (Z_1, Z_2, ..., Z_n)$ với $Z_i \sim \mathcal{N}(0, 1)$ là các bnn độc lập.
Khi đó X có dạng $X^t = A.Z^t + \mu^t$ với $A.A^t = \Sigma$. Suy ra

$$\Phi_X(s) = \mathbb{E}(e^{isX^t}) = \mathbb{E}(e^{i(sAZ^t + s\mu^t)}) = \mathbb{E}(e^{isAZ^t})e^{i\mu s^t}$$

Ta có $sAz^t = \sum_j(\sum_k s_k a_{kj})Z_j$ suy ra

$$\begin{aligned}\mathbb{E}(e^{isAZ^t}) &= \prod_j \mathbb{E}e^{i(\sum_k s_k a_{kj})Z_j} = \prod_j e^{-\frac{1}{2}(\sum_k s_k a_{kj})^2} \\ &= \prod_j e^{-\frac{1}{2}(\sum_{k,l} s_k a_{kj} s_l a_{lj})} = e^{-\frac{1}{2}\sum_k \sum_k s_k s_l (\sum_j a_{kj} a_{lj})} \\ &= e^{-\frac{1}{2}s\Sigma s^t}\end{aligned}$$

Vậy $\Phi_X(s) = e^{i\mu s^t - \frac{1}{2}s\Sigma s^t}$.

Bài tập 3.20. Xét phân phối của $\frac{X^2}{2}$

$$P(\frac{X^2}{2} < t) = P(-\sqrt{2t} < X < \sqrt{2t}) = \frac{2}{\sqrt{2\pi}}\int_0^{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Đạo hàm theo t

$$F'_{\frac{X^2}{2}}(t) = \frac{e^{-t}}{\sqrt{\pi t}} \Rightarrow \rho_{\frac{X^2}{2}}(t) = \frac{e^{-t}}{\sqrt{\pi t}} 1_{(0,+\infty)}(t)$$

Ta cũng có

$$\rho_{\frac{Y^2}{2}}(z) = \frac{e^{-z}}{\sqrt{\pi z}} 1_{(0,+\infty)}(z)$$

Đặt $X' = \frac{X^2}{2}, Y' = \frac{Y^2}{2}$ suy ra X', Y' là các bnn độc lập.
Ta có

$$\rho_{X',Y'}(t,z) = \frac{e^{-t-z}}{\pi\sqrt{tz}}1_{(t>0,z>0)}$$

Suy ra

$$\rho_{X'+Y'}(u) = \int_R \rho_{X'}(u-z)\rho_Y(z)dz = \frac{e^{-u}}{\pi}\int_0^u \frac{dz}{\sqrt{z(u-z)}} = e^{-u}$$

khi $u \geq 0$. ($\rho_{X'+Y'}(u) = 0$ khi $u < 0$)

Bài tập 3.21. Đặt $Z = (Z_1, Z_2) \Rightarrow \rho_Z(z_1, z_2) = \frac{1}{2\pi}e^{-\frac{z_1^2}{2}-\frac{z_2^2}{2}}$.
Xét $f(Z) = (\frac{Z_1}{Z_2}, Z_2)$, ta có $\rho_{f(Z)}(z_1, z_2) = \frac{|z_2|}{2\pi}e^{-\frac{z_1^2}{2}-\frac{z_2^2}{2}}$.
Với $a = \frac{z_1}{z_2}, b = z_2 \Rightarrow z_1 = ab$. Suy ra $\rho_{f(Z)}(a, b) = \frac{|b|}{2\pi}e^{-(a^2+1)\frac{b^2}{2}}$.
Ta có:

$$\begin{aligned}\rho_{\frac{Z_1}{Z_2}}(a) &= \int_R \rho_{f(Z)}(a,b)db = \frac{1}{2\pi}\int_{-\infty}^{+\infty}|b|e^{-(a^2+1)\frac{b^2}{2}} \\ &= \frac{1}{\pi(a^2+1)}\end{aligned}$$

Bài tập 3.22. i) Đặt $U = X, V = Z$ suy ra $X = U, Y = V\sqrt{1-r^2} + rU$.
Ta có:

$$\begin{aligned}\rho_{U,V}(u,v) &= \rho_{X,Y}(x,y)\sqrt{1-r^2} = \rho_{X,Y}(u, v\sqrt{1-r^2}+ru)\frac{1}{\sqrt{1-r^2}} \\ &= \frac{1}{2\pi}e^{-\frac{u^2+v^2}{2}}\end{aligned}$$

Suy ra $\rho_U(u) = \int_R \rho_{U,V}dv = \frac{1}{2\pi}e^{-\frac{u^2}{2}}$ và ta cũng có $\rho_V(v) = \frac{1}{2\pi}e^{-\frac{v^2}{2}}$, $\rho_U.\rho_V = \rho_{U,V}$ Vậy X, Z là các bnn độc lập và có phân bố $\mathcal{N}(0,1)$.

ii) Sử dụng phép đặt ở câu trên. Ta có

$$P(X>0,Y>0)=\int\limits_0^{+\infty}\int\limits_0^{+\infty}\rho_{X,Y}(x,y)dxdy$$

$$=\int\limits_{u>0}\int\limits_{v>\frac{-ru}{\sqrt{1-r^2}}}\rho_{U,V}(u,v)dudv$$

$$=\int\limits_{u>0}\int\limits_{v>0}\rho_{U,V}dudv+\int\limits_{u>0}\int\limits_{0>v>\frac{-ru}{\sqrt{1-r^2}}}\rho_{U,V}dudv=\frac{1}{4}+I(r)$$

Áp dụng công thức đạo hàm của tích phân ta có

$$I'(r)=\int\limits_{u>0}(\int\limits_{0>v>\frac{-ru}{\sqrt{1-r^2}}}\rho_{U,V}dv)'_r du$$

$$=\frac{1}{2\pi}\int\limits_0^{+\infty}(e^{-\frac{u^2}{2(1-r^2)}}\frac{1}{\sqrt{1-r^2}})du=\frac{1}{2\pi\sqrt{1-r^2}}$$

Suy ra $I=\frac{1}{2\pi}arcsin(r)\Rightarrow P(X>0,Y>0)=\frac{1}{4}+\frac{1}{2\pi}arcsin(r)$.

iii) Sử dụng phép đặt ở câu i). Ta có $Y=\sqrt{1-r^2}Z+rX$.
Vì Z,X là các bnn độc lập có phân phối chuẩn tắc nên dễ thấy Y cũng là phân phối chuẩn tắc.
Ta có:

$$\rho_{X|Y}(x\mid y)=\frac{\rho_{X,Y}(x,y)}{\rho_Y(y)}=\frac{1}{\sqrt{2\pi(1-r^2)}}exp(-\frac{(x-ry)^2}{2(1-r^2)})$$

Suy ra $P_{X|Y=y}$ là phân bố chuẩn có kỳ vọng là ry và phương sai là $1-r^2$ (không phụ thuộc vào y).

1.4 Lời giải bài tập Chương 4

Bài tập 4.1. Kí hiệu X_n là biến nhị phân được xác định như sau:

$$X_n = \begin{cases} 1 \text{ nếu lần tung thứ } n \text{ xuất hiện mặt 6,} \\ 0 \text{ trong trường hợp ngược lại.} \end{cases}$$

Khi đó có thể xem $\{X_n\}_{n\geq 0}$ là dãy phép thử Bernoulli với

$$p = P(X_n = 1) = \frac{1}{6}.$$

Kí hiệu

$$Z_n = \frac{\sum_{k=1}^{n} X_k - pn}{\sqrt{np(p-1)}}.$$

Với $n = 6000$ là một số nguyên dương đủ lớn, theo định lý Moivre-Laplace ta có:

$$\begin{aligned} &P(850 \leq S_{6000} \leq 1050) = \\ &\qquad = P\left(\frac{850 - 6000\cdot\frac{1}{6}}{\sqrt{6000\cdot\frac{1}{6}\cdot\frac{5}{6}}} \leq Z_{6000} \leq \frac{1050 - 6000\cdot\frac{1}{6}}{\sqrt{6000\cdot\frac{1}{6}\cdot\frac{5}{6}}}\right) \\ &\qquad\qquad\qquad\qquad \approx \phi(\sqrt{3}) - \phi(-3\sqrt{3}), \end{aligned}$$

trong đó ϕ là ký hiệu làm phân phối xác suất của phân bố normal chuẩn tắc $\mathcal{N}(0,1)$. (Bạn đọc tự tính toán tiếp!)

Bài tập 4.2. Giải tương tự như bài tập 4.1.

Bài tập 4.3. a) Ta biết rằng phân bố Poisson với tham số λ (kí hiệu là $P(\lambda)$)có hàm đặc trưng là:

$$\phi_\lambda(t) = e^{\lambda(e^{it}-1)}.$$

Vậy nếu $S_n = \sum_{k=1}^n X_k$ là tổng các biến ngẫu nhiên độc lập có cùng phân bố $P(1)$ thì hàm đặc trưng của S_n là:

$$\phi_{S_n}(t) = [\phi_{X_1}(t)]^n = e^{n(e^{it}-1)}.$$

Điều này chứng tỏ S_n có phân bố Poisson với tham số $\lambda = n$.

b) Bởi câu a, ta có thể xem $X_n = \sum\limits_{k=1}^{n} \xi_k$ với $\{\xi_k\}_{k\geq 1}$ là dãy biến ngẫu nhiên độc lập với cùng phân bố $P(1)$. Áp dụng định lý giới hạn trung tâm ta có:

$$P(X_n \leq n) = P\left(\frac{\sum_{k=1}^n \xi_k - n}{\sqrt{n}} \leq 0\right) \xrightarrow{n\to\infty} \int_{-\infty}^{0} \frac{e^{-\dfrac{x^2}{2}}}{\sqrt{2\pi}} = \frac{1}{2}.$$

Tuy nhiên ta cũng có thể viết lại $P(X_n \leq n)$ theo một cách khác:

$$P(X_n \leq n) = \sum_{k=0}^{n} P(X_n = k) = \sum_{k=0}^{n} e^{-n}\frac{n^k}{k!}.$$

So sánh hai hệ thức trên ta có điều cần chứng minh.

Bài tập 4.4. Theo định lý về tính liên tục ta có: Dãy biến ngẫu nhiên $\{X_n\}$ với phân bố $\mathcal{N}(\mu_n, \sigma_n^2)$ hội tụ yếu đến bnn X

$$\begin{aligned}
&\Leftrightarrow \phi_{X_n} \longrightarrow \phi_X\\
&\Leftrightarrow \exp(i\mu_n t - \frac{\sigma_n^2 t^2}{2}) \longrightarrow \phi_X\\
&\Leftrightarrow i\mu_n t - \frac{\sigma_n^2 t^2}{2} \longrightarrow \ln\phi_X\\
&\Leftrightarrow (\mu_n, \sigma_n) \longrightarrow (\mu, \sigma).
\end{aligned}$$

với $\mu, \sigma \geq 0$ nào đó. Ta có điều cần chứng minh.

Bài tập 4.6. Trước hết chú ý rằng phân bố hình học với tham số p có hàm đặc trưng được tính như sau:

$$\phi_p(t) = \sum_{k\geq 1} e^{itk} p_k = \sum_{k\geq 1} e^{itk} p(1-p)^{k-1}$$
$$= \frac{p}{1-p} \sum_{k\geq 1} [(1-p)e^{it}]^k = \frac{pe^{ik}}{(1-e^{ik}) + pe^{ik}}$$

Vậy

$$\phi_{\frac{X_n}{n}}(t) = \phi_{X_n}(\frac{t}{n}) = \left(it.e^{-\frac{it}{n}} \cdot \frac{1-e^{\frac{it}{n}}}{\frac{it}{n}} + 1 \right)^{-1}.$$

Chuyển qua giới hạn ta được:

$$\lim_{n\to\infty} \phi_{\frac{X_n}{n}}(t) = \frac{1}{1-it}.$$

Vế phải chính là hàm đặc trưng của bnn có phân bố mũ với tham số $\lambda = 1$. Do đó $\dfrac{X_n}{n} \xrightarrow{w} \mathcal{P}(1)$. Do tính liên tục của phân bố mũ ta có điều phải chứng minh.

Bài tập 4.7. Bằng cách tính toán trực tiếp ta sẽ chỉ ra phân bố của

Y_n hội tụ đến phân bố mũ với tham số $\lambda = 1$. Thật vậy:

$$\begin{aligned}
F_{Y_n}(x) = P(Y_n \le x) &= P(\max_{1\le i\le n} X_i \ge 1 - \frac{x}{n}) \\
&= 1 - P(\max_{1\le i\le n} X_i \le 1 - \frac{x}{n}) \\
&= 1 - \prod_{i=1}^{n} P(X_i \le 1 - \frac{x}{n}) \\
&= \begin{cases} 1 - (1 - \frac{x}{n})^n \text{ nếu } x \ge 0, n >> 1, \\ 0 \text{ nếu } x < 0. \end{cases} \\
&\xrightarrow{n\to\infty} \begin{cases} 1 - e^{-x} \text{ nếu } x \ge 0, \\ 0 \text{ nếu } x < 0. \end{cases}
\end{aligned}$$

Vậy $Y_n \xrightarrow{d} \mathbb{P}$.

Bài tập 4.9. Theo định lý Fubini ta có:

$$\begin{aligned}
|\frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} \phi_X(s)ds| &= |\frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon}(\int_{\mathbb{R}} e^{isx}dP_X)ds| \\
&= |\frac{1}{2\varepsilon}\int_{\mathbb{R}}(\frac{1}{ix}e^{isx}|_{-\varepsilon}^{\varepsilon})dP_X| \\
&= |\int_{\mathbb{R}} \frac{\sin\varepsilon x}{\varepsilon x}dP_X| \\
&\le \int_{|x|\le\frac{\varepsilon}{2}} |\frac{\sin\varepsilon x}{\varepsilon x}|dP_X + \int_{|x|>\frac{\varepsilon}{2}} |\frac{\sin\varepsilon x}{\varepsilon x}|dP_X \\
&\le \int_{|x|\le\frac{\varepsilon}{2}} dP_X + \frac{1}{2}\cdot\int_{|x|>\frac{\varepsilon}{2}} dP_X \\
&= \frac{1}{2}(1 + P[\frac{-2}{\varepsilon}, \frac{2}{\varepsilon}])
\end{aligned}$$

Từ đó ta có

$$P[\frac{-2}{\varepsilon}, \frac{2}{\varepsilon}] \geq |\frac{1}{\varepsilon}\int_{-\varepsilon}^{\varepsilon} \phi_X(s)ds| - 1.$$

Phụ lục B

Phần mềm máy tính cho xác suất thống kê

Có hàng trăm phần mềm máy tính cho tính toán xác suất thống kê. Có thể chia chúng theo các tính chất sau:

- Phần mềm cho toán nói chung, với các chức năng tính toán xác suất thống kê (ví dụ như MAPLE, MATLAB), hay là phần mềm chuyên về xác suất thống kê (ví dụ như MINITAB, S-PLUS, SAS, SPSS), hay là chuyên dụng hơn nữa (để dùng trong một lĩnh vực hẹp có cần đến thống kê).

- Phần mềm phải trả tiền (ví dụ như các phần mềm vừa kể trên), hay là miễn phí (ví dụ như R).

- Độ mạnh, độ đầy đủ của các chức năng, và độ dễ sử dụng, v.v.

Mỗi chương trình có những điểm mạnh và điểm yếu, thích hợp với những đối tượng khác nhau. Ví dụ:

- MATLAB thích hợp cho những người cần tính toán hình thức nói chung và có thể dùng đến thống kê. Gần đây MATLAB cũng được giới tài chính chuyên nghiệp dùng nhiều trong các công việc tính toán, làm mô hình, simulation, v.v.

- Đối với những người dùng nhiều đến hồi qui và “data mining” (đào số liệu để tìm thông tin), thì những chương trình như SPSS có thể thích hợp hơn.

- Chương trình R là một chương trình thống kê miễn phí, mã mở, và rất mạnh, được nhiều người dùng, đặc biệt trong giới hàn lâm. R trước kia có điểm dở là khó sử dụng, nhưng ngày nay, cùng với sự phát triển của giao diện trực giác, đã trở nên dễ sử dụng hơn nhiều.

Các chương trình về cơ bản có nhiều nguyên tắc chung giống nhau, nên nếu đã sử dụng thành thạo một chương trình thì sẽ không quá khó khăn chuyển sang dùng chương trình khác.

Để tính toán những ví dụ thống kê trong quyển sách này, các tác giả dùng một chương trình tương đối gọn nhẹ (bù lại chỉ có ít chức năng) có tên là **gretl** (viết tắt từ: Gnu Regression, Econometrics and Time-series Library), một phần mềm thống kê mã mở do Allin Contrell (GS kinh tế Đại học Wake Forest) khởi xướng và nhiều người ủng hộ xây dựng. Chương trình này được nhiều người khen là rất thích hợp cho giảng dạy ở đại học. Một số ưu điểm của gretl là:
- Miễn phí, mã mở,
- Có giao diện trực giác, rất dễ sử dụng,
- Chạy trên nhiều hệ điều hành khác nhau,
- Thích hợp về dạng số liệu với các chương trình thông dụng khác,
- Có cộng đồng người sử dụng và người lập trình phát triển nhanh,

- Có thể nạp các số liệu thống kê từ nguồn bên ngoài về qua internet,
- Có các chức năng tính toán thống kê đủ mạnh, đủ dùng cho các sinh viên học về kinh tế lượng, cũng như cho môn xác suất thống kê ở bậc đại học.

Có thể tìm hiểu về gretl trên trang web: http://gretl.sourceforge.net/.

Phụ lục C

Bảng phân bố Z

Phân bố normal chuẩn tắc $\mathcal{N}(0,1)$ còn được gọi là phân bố Z.
Bảng phía trước cho xác suất $P_{\mathcal{N}(0,1)}(]-\infty, Z[)$.
Bảng phía sau cho xác suất đuôi $P_{\mathcal{N}(0,1)}(]Z, \infty[)$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177

1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Z	P{Z to oo}	Z	P{Z to oo}	Z	P{Z to oo}	Z	P{Z to oo}
2.0	0.02275	3.0	0.001350	4.0	0.00003167	5.0	2.867 E-7
2.1	0.01786	3.1	0.0009676	4.1	0.00002066	5.5	1.899 E-8
2.2	0.01390	3.2	0.0006871	4.2	0.00001335	6.0	9.866 E-10
2.3	0.01072	3.3	0.0004834	4.3	0.00000854	6.5	4.016 E-11
2.4	0.00820	3.4	0.0003369	4.4	0.000005413	7.0	1.280 E-12
2.5	0.00621	3.5	0.0002326	4.5	0.000003398	7.5	3.191 E-14
2.6	0.004661	3.6	0.0001591	4.6	0.000002112	8.0	6.221 E-16
2.7	0.003467	3.7	0.0001078	4.7	0.000001300	8.5	9.480 E-18
2.8	0.002555	3.8	0.00007235	4.8	7.933 E-7	9.0	1.129 E-19
2.9	0.001866	3.9	0.00004810	4.9	4.792 E-7	9.5	1.049 E-21

Tài liệu tham khảo

[1] F. M. Dekking, C. Kraaikamp, H. P. Lopuhaä, L. E. Meester, A modern introduction to probability and statistics – Understanding why and how, Springer, 2005.

[2] L. Gonick, W. Smith, The cartoon guide to statistics, HarperCollins Publishers, 1993.

[3] Ch. M.Grinstead, J. L. Snell, Introduction to probability, AMS, 1997.

[4] Darrel Huff, How to lie with statistics, 1954.

[5] L. B. Koralov, Ya. G. Sinai, Theory of probability and random processes, Universitext, 2nd edition, 2007.

[6] R. Meester, A natural introduction to probability theory, 2008.

[7] G. Shay, Introduction to probability with statistical applications, Birkhäuser, 2007.

[8] A. N. Shiryaev, Probability (Graduate texts in mathematics, Vol. 95), Springer, 1995.

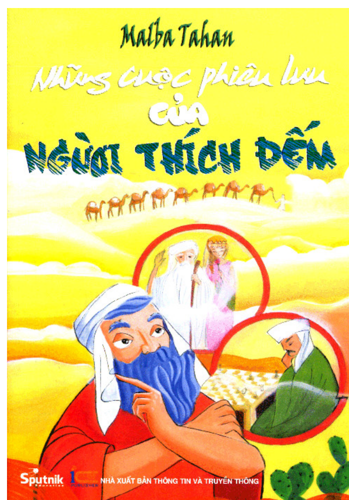
[9] Trần Mạnh Tuấn, Xác suất và thống kê – Lý thuyết và thực hành tính toán, NXB ĐHQGHN, 2004.

Giới thiệu Tủ Sách Sputnik

Các sách đã xuất bản

S001. Malba Tahan, Những cuộc phiêu lưu của Người Thích Đếm

Lê Hải Yến, Phạm Việt Hùng và Nguyễn Tiến Dũng dịch, 236 trang, 02/2015.



Đây là cuốn sách viết về toán học thường thức được ưa chuộng nhất trên thế giới trong vòng một thế kỷ qua. Nó đã được in ra hàng triệu bản, được dịch ra hầu hết các thứ tiếng phổ biến trên thế giới như tiếng Anh, tiếng Pháp, tiếng Tây Ban Nha, tiếng Đức, tiếng Ả Rập và được tái bản liên tục hàng năm...

Sự hấp dẫn đặc biệt của cuốn sách này nằm ở chỗ nó vừa là một

quyển sách giới thiệu rất nhiều điều thú vị về toán học, đồng thời vừa có giá trị rất cao về văn học và chứa nhiều điển tích lịch sử thú vị. Cuộc phiêu lưu của nhân vật chính trong cuốn sách ly kỳ không kém “Nghìn lẻ một đêm”.

Hợp với mọi lứa tuổi.

S002. Vladimir Levshin, Ba ngày ở nước Tí Hon

Nguyễn Tiến Dũng dịch, 190 trang, 02/2015



Đây là một quyển sách kỳ diệu, một “truyện thần thoại tuy không phải thần thoại” nhưng có phép màu làm cho cả học sinh và người lớn trở nên yêu toán học. Nó được nhà toán học Vladimir Levshin sáng tác ở Nga vào năm 1962 và từ đó đến nay được tái bản rất nhiều lần, tổng cộng hàng nghìn bản dịch, dịch sang các thứ tiếng khác nhau, và trở thành sách gối đầu giường của biết bao thế hệ học sinh. Ba ngày ở nước tí hon để lại ấn tượng sâu sắc trong hàng triệu bạn trẻ, và nhiều người trong số đó về sau sẽ trở thành nhà khoa học, kỹ sư, bác sĩ, thương gia, v.v.

Bản dịch của GS. Nguyễn Tiến Dũng do Sputnik xuất bản là bản dịch mới, chính xác hơn bản dịch cũ đã từng được in ở Việt Nam

trước đây.

Sách hợp với mọi lứa tuổi.

S003. Nguyễn Tiến Dũng, Các bài giảng về toán cho Mirella, quyển I

127 trang, 02/2015, kèm lời giới thiệu của GS. Hà Huy Khoái.

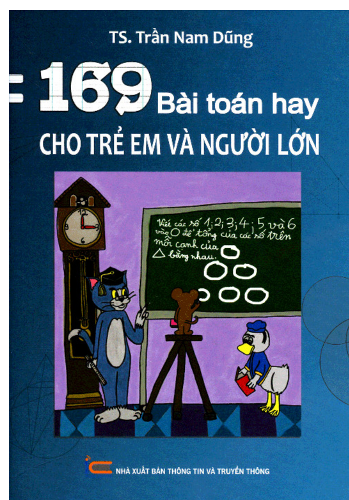


Cuốn sách gồm 12 chương, dựa trên các bài giảng và các buổi nói chuyện mà tác giả dành cho cô con gái của mình.

Trích từ một giới thiệu trong sách: Được viết bởi một nhà toán học hàng đầu là GS. TS Nguyễn Tiến Dũng , cuốn sách là một tài liệu quý và khác biệt gợi mở những vấn đề lý thú của toán học sơ cấp và hiện đại. Bản thân tôi rất ấn tượng với các bài giảng được dẫn dắt bằng ngôn ngữ gần gũi, hóm hỉnh nhưng rất logic và chứa đựng những ý tưởng sâu sắc của tác giả. Đây chắc hẳn là cuốn sách mà bất kỳ học sinh yêu toán nào cũng có thể tìm thấy những kiến thức bổ ích về toán học và việc học toán.

Sách dành cho các học PTCS và PTTH.

S004. Trần Nam Dũng, 169 bài toán hay cho trẻ em và người lớn

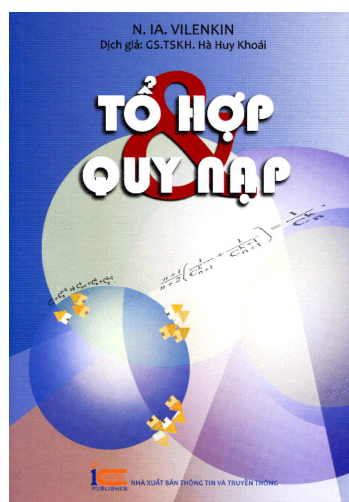


142 trang, 03/2015

Đây là cuốn sách bổ ích cho những bạn học sinh và những người yêu thích toán học. Với những bài toán được phát biểu rất vui, rất gần gũi trong cuộc sống, cuốn sách này sẽ đem lại cho các bạn những phút thư giãn cần thiết.

Sách hợp cho cả trẻ em và người lớn.

S005. N. Ia. Vilenkin, Qui nạp và tổ hợp



Hà Huy Khoái dịch, 03/2015, 87 trang.

Đây là một trong những cuốn sách viết hay và dễ hiểu nhất về phương pháp qui nạp và các vấn đề tính toán tổ hợp. Tác giả là nhà toán học Nga nổi tiếng N. Ia. Vilenkin. Sách hợp với trình độ phổ thông cơ sở và phổ thông trung học.

“Không chỉ quan trọng đối với những kỳ thi học sinh giỏi mà Tổ hợp và quy nạp là một phần không thể thiếu cho những ai muốn tiếp tục học tập, nghiên cứu và làm việc có hiệu quả trong những ngành toán học, tin học, kỹ thuật hay đơn giản chỉ là để trau dồi tư duy logic, điều mà ai cũng cần đến trong cuộc sống.”

Một số sách sắp xuất bản

Dưới đây là một số sách trong Tủ Sách Sputnik đã hoàn thàn hoặc hần như hoàn thành vào thời điểm 05/2015. Ngoài ra các cộng tác viên của Tủ Sách Sputnk đang viết và dịch nhiều quyển sách khác.

Lê Bích Phượng và Nguyễn Tiến Dũng, Romeo đi tìm công chúa, 100 câu đố vui hóc búa.

Quãng 140 trang, hoàn thành bản thảo 04/2015.

Sách này là một tuyển tập đúng 100 câu đố vui toán học, từ dễ đến khó, phù hợp với mọi lứa tuổi, chia thành các đề tài: số học, hình học, qui luật, thuật toán, và logic. Đặc biệt, có một chương về Romeo đi tìm công chúa, và để tìm được sẽ phải trải qua nhiều thử thách gian nan.

Ví dụ một câu đố từ quyển sách:

Romeo cùng với hai hiệp sĩ đi được đúng đường tới Động Tiên mà không bị sa vào bẫy. Bà tiên đã biết trước về sự xuất hiện của ba vị khách này, nên đã chuẩn bị sẵn 5 cái mũ, trong đó có hai cái màu xanh và ba cái màu đỏ. Bà tiên bảo ba chàng trai xếp thành 1 hàng

dọc, không được trao đổi với nhau, rồi đội lên đầu mỗi người một chiếc mũ từ năm chiếc đó. Romeo đứng đầu hàng, không nhìn thấy được bà tiên đội mũ màu gì lên đầu ai. Toto đứng giữa, nhìn thấy mũ trên đầu Romeo nhưng không nhìn thấy mũ trên đầu Dario và đầu mình. Dario đứng sau cùng nhìn thấy hai mũ trên đầu của Romeo và Toto nhưng không nhìn thấy mũ trên đầu mình. Bà tiên nói rằng "nếu ai suy luận được ra mũ mà mình đội màu nào một cách chắc chắn, thì sẽ được bà cho cái mũ đó". Cả Dario và Toto đều rất thông minh, nhưng đều lần lượt đành nói rằng họ không suy luận được mũ họ đội trên đầu màu gì. Đến lượt Romeo, thì Romeo lại suy luận được ra là đang đội mũ màu gì, và được bà tiên tặng cho cái mũ đó. Bạn có biết Romeo đội mũ màu gì không?

Nguyễn Tiến Dũng, Các bài giảng về toán cho Mirella, Quyển 2.

Quãng 170 trang, hoàn thành bản thảo 04/2015.

Có kèm lời giới thiệu của GS. Nguyễn Văn Mậu.

Tương tự như quyển "Mirella 1", mỗi chương sách của quyển này xuất phát từ một cuộc nói chuyện hay một bài giảng cho Mirella (con gái của tác giả) về toán học. Những vấn đề đề cập tới trong sách bao gồm: các đại lượng vô cùng nhỏ và vô cùng lớn, số học trên mặt phẳng, các hình đa diện lồi và các tính chất của chúng, đạo hàm và biến phân và ứng dụng của nó (ví dụ như định luật Snell trong quang học), các vấn đề về thuật toán và tin học, đặc biệt là khái niệm "lượng thông in", và các bài toán liên quan, ví dụ như bài sau:

Có 12 đồng tiền vàng trông giống hệt nhau, trong đó có 11 đồng tiền thật, và một đồng tiền giả. Các đồng tiền thật nặng bằng nhau, còn đồng tiền giả có khối lượng khác đồng tiền thật, nhưng không biết là nặng hơn hay nhẹ hơn. Dùng một cái cân cổ điển. Làm sao để với chỉ 3 lần cân mà chắc chắn xác định được rằng đâu là dồng tiền giả, và nó nhẹ hơn hay nặng hơn so với các đồng tiền thật.

Sách hợp với cuối cấp PTCS trở lên.

Lichtman, Bí mật, dối trá, và đại số.

Nguyễn Tiến Dũng dịch từ tiếng Anh, quãng 160 trang, hoàn thành bản thảo 03/2015.

Cuốn truyện cho thiên niên này đặc biệt ở chỗ nó có cốt truyện hẳn hoi, về cuộc sống và tình bạn của những học sinh lớp 8 ở một trường học ở Mỹ, đồng thời mỗi chương đều giới thiệu các ý tưởng và khái niệm toán học một cách rất tự nhiên và gần gũi cuộc sống. Cuốn sách nãy xuất bản bên Mỹ năm 2006, và đã đoạt nhiều giải thưởng về sách cho thiến niên.

Kiselev, Hình học phẳng.

Nguyễn Vân Hằng dịch từ tiếng Anh, quãng 360 trang, hoàn thành bản thảo 03/2015.

Đây là quyển sách kinh điển về hình học cho học sinh phổ thông, được dùng làm sách học chính thức ở Nga trong nhiều thập kỷ, và gần đây được dịch sang tiếng Anh. Nó trình bày một cách hệ thống và lô-

gích các khái niệm hình học phẳng (cho học sinh PTCS và PTTH), và kèm theo rất nhiều bài tập để qua đó học sinh có thể nắm chắc các kiến thức cơ bản.

Aleksandrova & Levshin, Người mặt nạ đen từ nươc Al-Jabr.

Nguyễn Tiến Dũng dịch từ tiếng Nga, quãng 240 trang, hoàn thành bản thảo 04/2015

Cuốn sách này cùng với hai cuốn sách khác là "Ba ngày ở nước Tí Hon" và "Thuyền trường Đơn Vị" (hay còn gọi là "Thủy thủ Số Không") tạo thành một bộ ba tập sách nổi tiếng do Levshin và Aleksandrova viết vào thập kỷ 1960. Từ đó đến nay, bộ sách này đã được tái bản liên tục hàng năm, in ra ở nhiều nước trên thế giới, trở thành "sách gối đầu giường" của hàng trăm nghìn bạn trẻ, những người mà về sau sẽ trở thành các nhà khoa học, bác sĩ, kỹ sư, thương gia, nhà quản lý, v.v.

Bản dịch mới này do Sputnik xuất bản tránh được nhiều lỗi sai của một bản dịch cũ đang được lưu hành tại Việt Nam.

Sách hợp với mọi lứa tuổi.

Đỗ Đức Thái, Các bài tập số học và đại số cho học sinh PTCS.

Vào năm 1993, GS. Đỗ Đức Thái (lúc đó là một tiến sĩ trẻ) có viết hai cuốn sách bài tập cho học sinh PTCS có năng khiếu về toán, một cuốn về số học và đại số, một cuốn về hình học. Hai cuốn sách đã đã

được in ra hàng chục nghìn bản, phục vụ nhiều thế hệ học sinh giỏi. Cuốn sách mới này của GS Đỗ Đức Thái là một phiên bản mở rộng và cập nhật của cuốn sách cũ đó về các bài tập đại số và số học.

Phiên bản mới của cuốn sách "Các bài hình học cho học sinh PTCS" của GS Đỗ Đức Thái cũng sẽ được Sputnik ra mắt bạn đọc trong tương lai gần.

Dành cho học các học sinh PTCS có năng khiếu về toán.

Spivak, Câu lạc bộ toán học lớp 6 - lớp 7.

Trần Nam Dũng dịch từ tiếng Nga.

Cuốn sách này, với khoảng 500 bài toán hay, có đi kèm lời giải, được xuất bản bằng tiếng Nga năm 2003, và được dùng làm sách học ở các lớp PTCS năng khiếu về toán ở Matskva.

Dành cho các học sinh PTCS.

Các địa chỉ bán sách Sputnik

Sách của Tủ sách Sputnik có được bán qua các công ty phát hành sách đến các cửa hàng online/offline các hội chợ sách, v.v. Ngoài ra, Sputnik có phân phối trực tiếp sách đến các địa điểm sau, những ai muốn mua có thể liên lạc. Danh sách này sẽ thỉnh thoảng được cập nhật.

Saigon (và khu vực miền Nam)

- Nhà sách Cá Chép, 211-213 Võ Văn Tần, TP HCM, 08 6290 6951

- Ms. Vũ Thị Bích Phương, Titan Education, 94 Mạc Đĩnh Chi, Mobile: 0909058520 Email: phuong@titan.edu.vn

- Titan Education (địa điểm khác), 175 Phạm Hùng, P. 4, Q. 8, TP HCM, Mobile: 0909058520 Email: phuong@titan.edu.vn

- Mr. Sơn, số điện thoại 0947558338 . Có trang FB Sách cho trẻ Sachchotre. Có thể giao sách tận nơi.

Đà Nẵng (và khu vực miền Trung)

- 111/18 Thanh Thủy, Đà Nẵng. Số điện thoại: 0906016943 hoặc 01667286280. (Có thể gọi điện, giao sách tận nhà nếu cần)

Hanoi và các nơi khác (có thể gọi điện hẹn lấy hoặc mua sách qua bưu điện)

- Booksquare (Quảng trường Sách), 12 Hòa Mã, Quận HBT, HN, Ms. Thủy, 04 3821 3888

- Trung tâm dạy toán Pomath, Ngõ 158 Nguyễn Khánh Toàn, Cầu Giấy, HN Ms. Hiên 091 513 7066

- Trung tâm CSVC và thiết bị, đồ chơi trẻ em – Viện KHGD VN 62 Phan Đình Giót, quận Thanh Xuân, HN Mrs. Cao Chi (84.4) 38642687

- Nhà sách Sư phạm, 12H1 Khu tập thể Đại học Sư phạm, Hà nội. Đt: 0437548642.

- Ms. Nguyễn Thị Thu – 241 phố Trần Đăng Ninh, Cầu Giấy Mobile: 0982932219 Email: sach@sputnik.vn (nhận gửi sách qua bưu điện)

- Ms. Quỳnh Anh, Ngõ 291 Lạc Long Quân, Nghĩa Đô, HN 093 518 5555 (nhận gửi sách qua bưu điện)

- Mrs Phượng, ngõ 43, đường Cổ Nhuế, 090 206 1246 (nhận gửi sách qua bưu điện)

- Mrs. Hà 090 200 8386 (nhận gửi sách qua bưu điện)

- Mrs. Thanh 091 323 9846 (liên hệ về phân phối sách, mua số lượng lớn)

Online

- Tiki.vn

Chỉ mục